Molecular adaptations and post-translational regulation of C$_4$-NADP-malic enzyme

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Anastasiia Bovdilova

From Kiew, Ukraine

Düsseldorf, January 2019

from the Institute of Developmental and Molecular Biology of Plants,
at the Heinrich Heine University Düsseldorf

Посвящается моим родителям с безграничной любовью и благодарностью

Table of contents

List of abbreviations

2-PG	2-phosphoglycolate
3-PGA	3-phosphoglycerate
Ala	Alanine
AlaAT	Alanine aminotransferase
Asp	Aspartate
AspAT	Aspartate aminotransferase
AT	Aminotransferase
ATP	Adenosine triphosphate
BS	Bundle sheath
C_3 acid	An organic acid molecule containing three carbon atoms
C_3 plants	Plant species performing C_3 photosynthesis
C_4 acid	An organic acid molecule containing four carbon atoms
C_4-NADP-ME	NADP-ME isoform involved in C_4 potosyntesis
C_4 plants	Plant species performing C_4 photosynthesis
CA	Carbonic anhydrase
CAM	Crassulacean acid metabolism
CCM	Carbon concentrating mechanism
EC	Enzyme Commission
e.g.	For example
et al.	And others (Latin: *Et alii*)
etc.	And so forth (Latin: *Et cetera*)
Gln	Glutamine
Glu	Glutamic acid
LC	Liquid chromatography
Leu	Leucine
M	Mesophyll
MDH	Malate dehydrogenase
ME	Malic enzyme
MS	Mass spectrometry
NAD	Nicotinamide adenine dinucleotide
NAD-ME	NAD-dependent malic enzyme

NADP	Nicotinamide adenine dinucleotide phosphate
NADP-ME	NADP-dependent malic enzyme
nonC$_4$-NADP-ME	Chloroplastic housekeeping NADP-ME isoform
NPR1	Nonexpressor of pathogenesis-related genes 1
OAA	Oxaloacetate
PCA	Primary carbon assimilation
PCR	Primary carbon reduction
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PEP-CK	Phosphoenolpyruvate carboxykinase
Phe	Phenylalanine
PPDK	Pyruvate orthophosphate dikinase
PS	Photosystem
PTM	Post-translational modifications
Pyr	Pyruvate
RubisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	D-ribulose-1,5-bisphosphate
SAR	Systemic acquired resistance
Ser	Serine
SWATH	Sequential window acquisition of all theoretical mass spectra
Thr	Threonine
TP	Triose phosphate
Tyr	Tyrosine
V	Vein

1 Summary

C_4 photosynthesis is a carbon-concentrating mechanism, which selectively enriches carbon dioxide around RubisCO and simultaneously reduces wasteful photorespiration. In plant species exhibiting the two-celled C_4 pathway, this selective CO_2 accumulation is achieved due to the spatial separation of the photosynthetic processes between mesophyll and bundle sheath cells. Depending on the type and subcellular localization of the predominant decarboxylase in the bundle sheath cells, plants performing C_4 photosynthesis can be divided into three different subtypes – NADP-ME, NAD-ME and PEP-CK variants. The photosynthetic NADP-ME isoform (C_4-NADP-ME) has evolved from its closely-related plastidic non-photosynthetic isoform (nonC_4-NADP-ME). In the course of evolution, C_4-NADP-ME has acquired several unique biochemical and structural properties, which are, together with the differences in the expression levels and patterns, crucial for fulfilling its new biological function.

In this thesis, main molecular events leading to the C_4-related biochemical and structural characteristics of the C_4-NADP-ME from maize were elucidated. An algorithm has been developed to identify amino acid residues differentially conserved in C_4 and nonC_4 isoforms of closely related C_4-NADP-ME grasses. Moreover, the crystal structure of the maize C_4-NADP-ME in its apoform was obtained. Based on these studies, four candidate amino acids were selected for deeper biochemical and structural analyses. Results of these experiments performed with the wild-type enzyme and mutagenized forms, together with the results of the crystallization study confirmed the involvement of the candidate amino acid residues in conferring fundamental C_4-properties: Phe140 is critical for the homotetrameric assembly of the C_4-NADP-ME; and Gln503, Leu544 and Glu339 are responsible for the high affinity to the substrate malate.

In parallel, mass spectrometry analyses identified a unique serine phosphorylation site within the C_4-NADP-ME in protein extracts of maize leaves. Further analysis revealed that phosphorylation of this serine increases during the first hours of illumination, with the highest levels observed at two hours into light. *In vitro* studies with the phosphomimetic C_4-NADP-ME showed reduced activity of the enzyme compared to the wild-type, implying that phosphorylation is a mechanism to reduce the decarboxylase activity *in vivo*. Analysis of the crystal structure of sorghum C_4-NADP-ME with bound $NADP^+$ indicated that the serine residue is involved in $NADP^+$ binding at the active site. This allowed to propose a molecular mechanism for the inhibitory influence of the phosphorylation event.

2 Zusammenfassung

Die C$_4$-Photosynthese stellt ein Mechanismus zur Kohlenstoff-Anreicherung dar, welcher selektiv Kohlenstoffdioxid in der Umgebung von RubisCO anreichert und gleichzeitig den Verlust von fixiertem Kohlenstoff während der Photorespiration verringert. In den Pflanzenarten, welche den auf zwei Zellarten-basierenden C$_4$-Stoffwechselweg nutzen, wird diese zielgerichtete Kohlenstoffdioxid-Akkumulation mittels der räumlichen Aufteilung der photosynthetischen Prozesse zwischen den Mesophyll- und Bündelscheidenzellen erreicht. Abhängig von der Art und der zellulären Lokalisierung der dominierenden Decarboxylase in den Bündelscheidenzellen, können die C$_4$-Pflanzen in drei unterschiedliche Untergruppen unterteilt werden – die NADP-ME, NAD-ME und die PEP-CK Varianten. Die photosynthetische NADP-ME Isoform (C$_4$-NADP-ME) hat sich aus ihrer eng verwandten nicht-photosynthetischen Isoform (nonC$_4$-NADP-ME) im Lauf der Evolution entwickelt. Die C$_4$-NADP-ME Isoform hat sich einige einzigartige biochemische und strukturelle Eigenschaften angeeignet, die, zusammen mit den Unterschieden in den Expressionsleveln und –mustern, entscheidend für das Erfüllen der neuen biologischen Funktion sind.

In dieser Doktorarbeit wurden die wichtigsten molekularen Veränderungen aufgeklärt, welche zu den C$_4$-bezogenen biochemischen und strukturellen Eigenschaften der C$_4$-NADP-ME von Mais im Laufe der Evolution geführt haben. Mit Hilfe eines Algorithmus konnten die unterschiedlich konservierten Aminosäurenreste in den C$_4$- und nonC$_4$-Isoformen von eng verwandten C$_4$-NADP-ME Gräsern identifiziert werden. Außerdem wurde die Kristallstruktur von C$_4$-NADP-ME aus Mais in ihrer Apoform ermittelt. Basierend auf diesen Studien, wurden vier Aminosäurenreste für die weiterführenden biochemischen und strukturellen Untersuchungen ausgewählt. Die Ergebnisse dieser Experimente, welche mit der wildtypischen und den mutagenisierten Enzymvarianten durchgeführt wurden, zusammen mit den Ergebnissen der Kristallisierungsstudie bestätigten die Beteiligung dieser vier Aminosäurenreste an der Ausprägung der wesentlichen C$_4$-Eigenschaften: Phe140 ist essentiell für die homotetramerische Anordnung von C$_4$-NADP-ME; und die Gln503, Leu544 und Glu339 sind verantwortlich für die hohe Affinität zum Substrat Malat.

Parallel dazu wurde eine Phosphorylierungsstelle an einem Serinrest in dem Mais C$_4$-NADP-ME mittels massenspektrometrischen Analysen identifiziert. Weiterführende Versuche zeigten, dass diese Serinphosphorylierung von C$_4$-NADP-ME in den ersten Stunden der Belichtung zunahm; dabei wurde die höchste Phosphorylierung nach zwei Stunden Lichtexposition beobachtet. *In vitro* Studien mit der heterologen Wildtyp- und der

mutagenisierten Enzymform mit dem nachgeahmten Phosphorylierungszustand zeigten eine verringerte kinetische Leistung des mutagenisierten Enzyms und implizieren, dass die Phosphorylierung ein Mechanismus zur Verringerung der Aktivität dieser Decarboxylase *in vivo* darstellt. Die Analyse der Kristallstruktur der Sorghum C4-NADP-ME mit dem gebundenen NADP$^+$ weist auf die Beteiligung dieses Serinrestes bei der NADP$^+$-Bindung im aktiven Zentrum hin. Das liefert einen Hinweis auf den molekularen Mechanismus für den hemmenden Einfluss der Phosphorylierung.

3 Introduction

3.1 C_4 photosynthesis represents a unique recent evolutionary acquisition of the angiosperms

More than 50 years ago, Hatch and Slack showed that malate, aspartate, and oxaloacetate contained more than 93% of radioactive $^{14}CO_2$ introduced to detached sugarcane leaves within the first second of exposure; they proposed a biochemical pathway explaining this phenomenon (Hatch and Slack, 1966). Identified organic acid molecules contain four carbon atoms and are denominated C_4 acids. Plant species performing a variation of photosynthesis in which CO_2 from air is first incorporated into the C_4 acids have therefore been collectively designated C_4 plants (Sage, 2004; Ludwig, 2013). In contrast to this, C_3 plants perform an evolutionarily more ancient kind of photosynthesis. In this case, CO_2 is directly fixed by ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) on D-ribulose-1,5-bisphosphate (RuBP), which results in two molecules of a three-carbon acid 3-phosphoglycerate (3-PGA) (Wang et al., 2011).

This carboxylation reaction is the first enzymatic rate-limiting step in the photosynthetic carbon reduction (Calvin) cycle (Raines, 2003). In addition to catalyzing the life-sustaining carboxylation reaction (Erb and Zarzycki, 2018), this enzyme has two characteristic drawbacks: a remarkably low catalytic turnover frequency and its ability to accept molecular oxygen as substrate, which leads to an additional oxygenation reaction of RuBP (Andersson, 2008; Gruber and Feiz, 2018). The oxygenation reaction results in a production of one 2-phosphoglycolate (2-PG) molecule in addition to a 3-PGA molecule. This catalytic "error" can be explained by evolution of RubisCO, which started more than 3.5 billion years ago prior to the oxygenation of the Earth atmosphere, when there was no need for the enzyme's active site to distinguish between these two gas substrates (Whitney et al., 2011; Erb and Zarzycki, 2018).

3.1.1 C_4 plants are characterized by distinct biochemical and anatomical features

C_4 plants differ from the plants exhibiting C_3 photosynthesis not only by the product of the initial atmospheric CO_2 fixation, but also by several other factors, the most important of which are the spatial and biochemical organization of the photosynthetic process and changes in the ultrastructure of leaves (Sage and Zhu, 2011).

In C_3 plants primary carbon acquisition (PCA) and primary carbon reduction (PCR) occur within a single photosynthetic cell, whereas in C_4 plants PCA and PCR take place in different photosynthetic cell-types (Langdale and Nelson, 1991). These photosynthetic cell types in C_4 plants are mesophyll (M) and bundle sheath (BS) cells. Exceptions are plants exhibiting the

obligate and facultative single-cell C_4 photosynthesis, since in these species carbon assimilation takes place within the same cell (von Caemmerer et al., 2014). However, also in these cases the spatial separation is achieved through PCA and PCR being localized to different cellular compartments (Langdale, 2011; von Caemmerer et al., 2014; Sharpe and Offermann, 2014).

In most terrestrial C_4 plants, the spatial division of photosynthetic processes was found to be associated with anatomical diversification of the leaf structure, the Kranz anatomy (Langdale, 2011). In a classical example of Kranz anatomy, veins (V) are surrounded by a layer of enlarged BS cells, which are in turn surrounded by a concentric layer of M cells (Hatch, 1987; Ludwig, 2013; Fouracre et al., 2014). Each set of two veins are therefore separated by two BS and two M cells leading to a characteristic V-BS-M-M-BS-V wreath-like pattern (Gowik and Westhoff, 2011; Langdale, 2011). In comparison, C_3 plants usually have much smaller BS cells, which are separated by a higher number of up to 20 M cells (Langdale, 2011; Ludwig, 2013), which leads to increased inter-veinal distances and decreased BS:M ratios compared to C_4 plants (Nelson, 2011; Griffiths et al., 2013). Additionally, the Kranz syndrome is often associated with specific changes on the ultrastructural level of C_4 leaf organization (Hatch, 1987). These include suberization of the cell walls of BS cells, which occurs in several C_4 grass species (Mertz and Brutnell, 2014); formation of fewer chloroplasts in the M cells than in closely related C_3 species (Stata et al., 2014) and enriching plasmodesmata density between the M and BS cells (Botha, 1992; Sage et al., 2014; Danila et al., 2016).

In all plants performing the two-celled C_4 pathway, PCA is enabled by means of the cytoplasmic phosphoenolpyruvate carboxylase (PEPC), which is localized in M cells and lacks oxygenase activity (Andreo et al., 1987; Hatch, 1987). Cytosolic carbonic anhydrase (CA) is responsible for the first enzymatic step of the PCA, which reversibly hydrates the atmospheric CO_2 (Hatch and Burnell, 1990). Subsequently, PEPC assimilates generated bicarbonate ions, leading to the production of the first C_4 acid, oxaloacetate (OAA) (Maier et al., 2011). OAA is then immediately converted to the more stable C_4 acid, malate and/or aspartate, which subsequently moves to the neighboring BS cells, where it is decarboxylated in the direct proximity of RubisCO, which selectively increases the concentration of CO_2 around it leading to the decrease in its oxygenation and photorespiration rates (Keeley and Rundel, 2003; Ludwig, 2013; Moroney et al., 2013). This biochemical feature is known as carbon concentrating mechanism (CCM) and makes C_4 photosynthesis the most efficient form of photosynthesis on Earth (Sage and Zhu, 2011). Based on the type and subcellular localization of the enzyme which predominantly decarboxylases C_4 acids in the BS cells, C_4 plants have

been classified into three different subtypes: nicotinamide adenine dinucleotide phosphate (NADP)-dependent malic enzyme (NADP-ME), nicotinamide adenine dinucleotide (NAD)-dependent malic enzyme (NAD-ME) and adenosine triphosphate (ATP)-dependent phosphoenolpyruvate carboxykinase (PEP-CK) (Maier et al., 2011; Ludwig, 2016). However, accumulationg data indicating simultaneous action of several decarboxylases within BS cells of many C_4 plant species suggest that classification into only three biochemical subtypes is oversimplifying (Furbank, 2011; Furbank, 2017). Apart from CO_2, another product of the decarboxylation reaction is a C_3 organic acid in the form of pyruvate or phosphoenolpyruvate, which is then returned to the M cells, where the CO_2 acceptor molecule phosphoenolpyruvate (PEP) is regenerated (Hatch, 1987).

Evolution of C_4 plants from their C_3 ancestors began quite recently in geological time, about 30-35 million years ago (Edwards et al., 2010; Sage and Stata, 2015). Emergence of this trait can be linked to drastic changes in the climate which occurred around the same time and were characterized by decrease in the CO_2 levels in the atmosphere (Christin et al., 2008). Under these changing environmental conditions the inability of RubisCO to distinguish between CO_2 and O_2 became critical, since a product of the oxygenation reaction, 2-PG, is toxic for plant metabolism (Anderson, 1971; Bauwe et al., 2010). In order to detoxify this compound, the photorespiratory pathway has evolved. However, it is energy-consuming and is combined with losses of previously fixed carbon (Zhu et al., 2008; Peterhansel and Maurino, 2011; Moroney et al., 2013). Evolutionary acquirement of the C_4 trait enabled significant physiological benefits, such as increased water and nitrogen use efficiencies in comparison to the ecologically similar C_3 plant species growing in arid, high light or saline environments (Langdale, 2011; Sage and Zhu, 2011).

Evolutionary analyses of the C_4 pathway development have indicated that it has evolved independently between 65–84 times in flowering plants (Sage, 2016). This makes this trait a demonstrative example of the convergent evolution of a complex trait (Bräutigam and Gowik, 2016). From these origins ~8100 of both mono- and dicotyledonous C_4 species belonging to 19 families have arisen (Sage, 2016). Even though this number of species only makes up about 3% of the total species number of present-day angiosperms (Sage and Stata, 2015), C_4 plants are estimated to contribute about a quarter of the total net productivity in the terrestrial biosphere (Still et al., 2003).

3.1.2 C$_4$ pathway in maize operates based on parallel production and utilization of two C$_4$ acids

Maize (*Zea mays*) is the world's most widely grown C$_4$ crop (Leakey, 2009; Sage, 2016) with a paramount historical, social, scientific and economical importance (Covshoff et al., 2014; Sage, 2016). The first step of the C$_4$ cycle operating in maize is the same as in all dual-cell C$_4$ plants, and is characterized by the CA-catalyzed conversion of the atmospheric CO$_2$ into bicarbonate, which occurs in the cytoplasm of M cells (Maier et al., 2011; Ludwig, 2016; Rao and Dixon, 2016) (Figure 1, solid black arrows). This reaction is followed by assimilation of bicarbonate via PEPC, leading to OAA production. OAA is then imported into the chloroplasts of the M cells, where it is reduced to malate by NADP-dependent malate dehydrogenase (MDH) (Edwards and Nakamoto, 1985). Malate diffuses into the BS cells along its concentration gradient (Arrivault et al., 2017), where it is transported into the chloroplasts and is immediately decarboxylated by NADP-ME, resulting in production of one molecule each of NADPH, CO$_2$ and pyruvate (Maier et al., 2011; Rao and Dixon, 2016). Released CO$_2$ is directly fixed by RubisCO, which is also localized in the BS chloroplasts. Pyruvate is transferred to the M chloroplasts, where it is converted into PEP by action of pyruvate orthophosphate dikinase (PPDK) (Edwards and Nakamoto, 1985). Regenerated PEP is exported to the cytosol (Maier et al., 2011; Ludwig, 2016; Rao and Dixon, 2016) (Figure 1).

Evidence for the existence of an additional decarboxylation pathway, operating in parallel and based on the PEP-CK activity, has been accumulated over years (Figure 1, turquoise dashed arrows). A metabolic hint came from early experiments with ^{14}CO$_2$, which were performed with photosynthesizing maize leaves. Results of these studies showed that even though the highest amount of total radiolabel could be detected in malate, a smaller, but significant part (approximately 25%) was shown to be incorporated into aspartate (Asp) (Hatch, 1971). Furthermore, isolated strands of BS cells from maize were shown to possess a significant capacity to perform aspartate decarboxylation (Chapman and Hatch, 1981). In addition, presence of PEP-CK exclusively in the BS maize cells was shown with immunolocalization studies, and activity of the decarboxylase was measured (Walker et al., 1997). More recently, Pick and collaborators revised the model of the C$_4$ cycle in this monocot plant species (Pick et al., 2011). According to this model, after import into M chloroplasts, OAA can be converted not only into malate, but also into aspartate by means of the chloroplastic aspartate aminotransferase (AspAT) (Figure 1, turquoise dashed arrows). Aspartate is transferred into BS cells, where it is imported into chloroplasts and converted into OAA by AspAT. Resulting OAA has two possible fates: it may be converted to malate by chloroplastic MDH, or it can be

exported to the BS cytosol, where it serves as a substrate for decarboxylation by PEP-CK leading to PEP and CO_2 production (Pick et al., 2011). Additionally, successful functioning of the PEP-CK decarboxylation cycle is supported by another enzyme, alanine aminotransferase (AlaAT). AlaAT is responsible for conversion of pyruvate to alanine in BS and M cells and participates in maintaining ammonia balance between these two cell types (Pick et al., 2011; Rao and Dixon, 2016) (Figure 1). In agreement with this, transcripts encoding maize AlaAT were shown to be evenly distributed between M and BS (Chang et al., 2012; Schlüter et al., 2016).
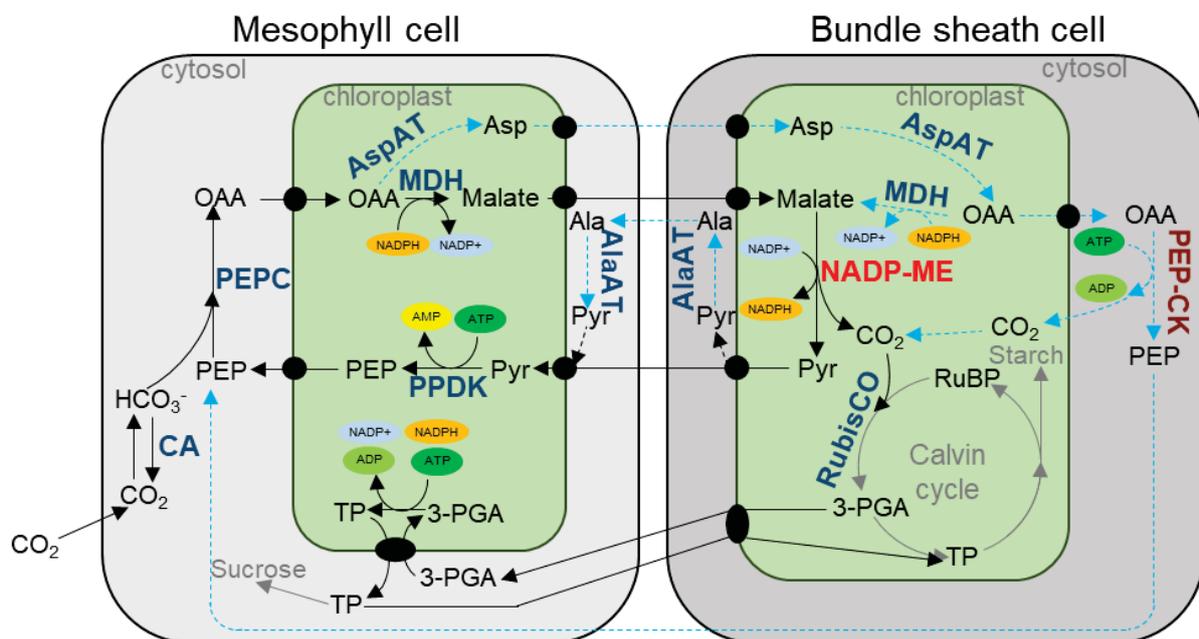
Figure 1: Schematic representation of C₄ pathway and carbon flow resulting from parallel operating of two decarboxylases in maize leaves. 3-PGA, 3-phosphoglycerate; ADP, adenosine diphosphate; AlaAT, alanine aminotransferase; AMP, adenosine monophosphate; AspAT, aspartate transaminase; ATP, adenosine triphosphate; CA, carbonic anhydrase; MDH, NADP-dependent malate dehydrogenase; NADP, nicotinamide adenine dinucleotide phosphate; NADP-ME, NADP-dependent malic enzyme; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PEPC, phosphoenolpyruvate carboxylase; PEP-CK, phosphoenolpyruvate carboxykinase; PPDK, pyruvate orthophosphate dikinase; Pyr, pyruvate; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO); RuBP, D-ribulose-1,5-bisphosphate; TP, triose phosphates. Pathways modified after Gowik and Westhoff (2011), Pick et al. (2011) and Schlüter et al. (2016), 3-PGA/TP shuttle modified after Furbank et al. (1985).

3.1.3 Coordination of two C₄ decarboxylases in maize is crucial but underlying mechanisms are not yet completely elucidated

An essential point of operating a C₄ cycle based on simultaneous activity of two decarboxylases is a different cell-type-specific energy balance, which results from transfer of malate and Asp (Furbank, 2011; Ludwig, 2016). Maize BS chloroplasts were found to possess fewer grana than chloroplasts of M cells (Gutierrez et al., 1974; Hatch et al., 1975), congruently with the fact

that only very limited photosystem (PS) II activity could be reported for this cell type (Woo et al., 1970; Leegood et al., 1983; Meierhoff and Westhoff, 1993). This deprivation of PSII function is beneficial for efficient C_4 cycle functioning because it limits O_2 production, which additionally prevents the oxygenation of RubisCO. However, it also means that the linear electron transport cannot operate, which restricts the production of reducing equivalents in the BS cells (Munekage, 2016). Reduction of the 3-PGA molecules to triose phosphates (TPs) in the Calvin cycle requires NADPH (Raines, 2003). Since malate carries a reducing equivalent and Asp does not (Figure 1), it was long suggested that transport of malate is energetically more advantageous. However, after significant accumulation of evidence that reducing equivalents for fueling the Calvin cycle in BS cells are mostly generated by means of the 3-PGA/TP shuttle functioning between BS and M cells (Majeran et al., 2005; Bräutigam et al., 2008; Majeran et al., 2008), this assumption could be discarded (Pick et al., 2011). In this shuttle, 3-PGA is transported from BS cells into M cells, where it is imported into chloroplasts followed by reduction to TPs (Figure 1). Some of the TP produced in M chloroplasts is transported back to BS chloroplasts, where it enters the Calvin cycle and participates at the RuBP regeneration (Figure 1) (Weber and von Caemmerer, 2010; Weber and Bräutigam, 2013; Ludwig, 2016). The rest of the TP produced in M cells is used for sucrose synthesis (Furbank et al., 1985; Lunn and Furbank, 1997) (Figure 1). In contrast to this, starch synthesis in maize mostly occurs in the BS cells (Downton and Hawker, 1973) (Figure 1).

After establishment of the C_4 cycle in maize as a branched rather than a linear cycle (Pick et al., 2011), further questions arose: how are carbon and energy flow coordinated in such a system, and what advantage does this combined system offer? Large pools of the C_4 and C_3 acids are distributed asymmetrically between M and BS cells in illuminated maize leaves (Leegood, 1985; Stitt and Heldt, 1985; Arrivault et al., 2017). In addition to their role in enabling carbon flow between these two compartments by generating concentration gradients (Pick et al., 2011; Kromdijk et al., 2014; Wang et al., 2014), it has been proposed that these pools of acids additionally act as a buffer system that can store and release ATP and reducing equivalents in response to rapid changes in light availability. This provides an additional physiological advantage for C_4 plants, since no comparable buffer system is known in C_3 plants (Stitt and Zhu, 2014). In agreement with this, transamination and the two decarboxylase system are important for regulation of ATP and NADPH supply in M and BS cells under variable light qualities (Bellasio and Griffiths, 2014a). Furthermore, functioning of the two decarboxylation systems simultaneously may potentially enable plasticity in response to a wider range of environmental variables. As an example, such stress factors as salinity and sade treatments

exhibited a different effect on the NADP-ME and PEP-CK activity, which provides evidence for flexibility of the maize decarboxylase system in BS cells in response to changing environmental conditions (Sharwood et al., 2014).

3.2 The emergence of the main C$_4$ enzymes was strongly facilitated by gene duplications during the preconditioning phase of C$_4$ photosynthesis evolution

The C$_4$ trait evolved in parallel in several lineages of C$_3$ ancestors and is absent in the vast majority of angiosperm families, which implies that C$_4$ evolution was tightly associated with specific characteristics, or preconditioning traits, present only in certain taxa (Sage, 2004; Sage et al., 2012; Christin and Osborne, 2013). During subsequent evolution, these genetic and physiological traits present in the C$_3$ forefather species needed to be subjected to readjustments in order to adapt to a new C$_4$-related function (Christin and Osborne, 2013). For deeper understanding of the C$_4$ trait development within these lineages, a model has been developed, which proposes five main phases of evolution of a C$_4$ plant from its C$_3$ ancestor (Sage et al., 2012). These phases are the preconditioning, evolution of Proto-Kranz anatomy, evolution of C$_2$ photosynthesis, establishment of the C$_4$ metabolic cycle and optimization (Sage et al., 2012). The first phase would have been tightly associated with specific biochemical and anatomical preconditioning traits: high vein density, the acquisition of the regulatory elements that could be modified in order to enable C$_4$ specific patterns of gene expression, gene duplication events and large genome size (Sage et al., 2012).

The ability to generate and maintain a sufficient number of redundant genes represents a critical preconditioning trait. These redundant genes result from the duplications of whole genome, genome segment or single genes (Monson, 2003; Sage, 2004; Gowik and Westhoff, 2011). An advantage provided by these duplication events can be explained by the fact that in some cases resulting gene copies may acquire a novel, beneficial function. Simultaneously, the plant retains the cellular function supported by the unmodified protein transcribed from the original gene. This outcome in the evolution of the duplicated genes is referred to as neofunctionalization (Lynch and Conery, 2000). In agreement with this evolutionary scenario, orthologs of the enzymes participating in CO$_2$-pump generation and maintenance within known C$_4$ pathways are present in C$_3$ species (Westhoff and Gowik, 2004; Aubry et al., 2011; Ludwig, 2013). Remarkably, in contrast to C$_4$ plants, these ancestral proteins do not participate directly in carbon fixation and assimilation processes in C$_3$ plants, but are mostly involved in replenishment of intermediates for different metabolic pathways, pathogen-related defense reactions and other cellular processes (Maier et al., 2011; Ludwig, 2013).

In the course of evolution, a minimum of three major changes must have taken place in order to adapt the ancestor C_3 gene to its C_4-related function (Westhoff and Gowik, 2004). First, effectiveness of the gene expression had to be increased, since most C_4 genes are characterized by higher expression rates than their non-C_4 counterparts (Williams et al., 2012). Second, C_4 isoforms needed to acquire organ-, cell-, and organelle-specific expression patterns (Monson, 2003). Exact molecular changes leading to these quantitative and qualitative differences in gene expression are diverse and vary depending on the affected gene (Hibberd and Covshoff, 2010; Gowik and Westhoff, 2011). Third, many C_4 isoforms are known to exhibit altered kinetic and regulatory behaviour in comparison to their C_3 progenitors; this implies that changes in the coding region of these genes were also necessary (Maier et al., 2011; Ludwig, 2013). These kinetic adjustments were necessary for enabling efficient functioning of newly evolved C_4 enzymes in the changed metabolic environment (Saigo et al., 2013).

3.3 Malic enzymes are widespread in nature due to the metabolically relevant reaction they catalyze

Malic enzyme catalyzes the reversible oxidative decarboxylation of malate with the concomitant reduction of NAD^+ or $NADP^+$ resulting in production of pyruvate and CO_2 (Chang and Tong, 2003). For efficient catalysis, a bivalent cation, Mg^{2+} or Mn^{2+}, is required. MEs are classified according to their cofactor specificity and the ability to decarboxylate OAA into groups with following Enzyme Commission (EC) numbers: EC 1.1.1.38 (NAD dependent, OAA decarboxylating), EC 1.1.1.39 (NAD dependent, non-OAA decarboxylating) and EC 1.1.1.40 (NADP dependent, OAA decarboxylating) (International Union of Biochemistry and Molecular Biology Nomenclature, https://www.qmul.ac.uk/sbcs/iubmb/nomenclature/).

Since both the products and the substrates of this reaction possess substantial metabolic significance, a plethora of different ME isoforms exists, which are widely distributed in nature and are found in animals (Chang and Tong, 2003), plants (Wedding, 1989), bacteria (Bologna et al., 2007), yeast, and fungi (Vorapreeda et al., 2013). In C_4 and Crassulacean acid metabolism (CAM) plants, this enzyme has an especially important function as it is involved in providing CO_2 for RubisCO: NADP-ME in case of the NADP-ME subtype and NAD-ME in case of the NAD-ME and PEP-CK subtypes (Maier et al., 2011). Similar to this, in CAM plants, also species mostly utilizing PEP-CK for CO_2 production, have been shown to possess some amounts of ME activity (Dittrich et al., 1973).

MEs are localized in several cellular compartments: cytoplasm and mitochondria in mammals (Chang and Tong, 2003), and cytosol (Detarsio et al., 2008), mitochondria (Artus and Edwards,

1985) and chloroplasts (Maurino et al., 1997) in plants. In terms of their structural organization, MEs represent a unique class of oxidative decarboxylases, since the topology of their polypeptide backbone is different from the other oxidative decarboxylases (Chang and Tong, 2003).

3.3.1 Plant families of NADP-ME encompass several isoforms with distinct biological functions

Similar to mammals, who possess NADP-MEs localized in different cellular compartments (cytosol and mitochondria) (Loeber et al., 1994a; Loeber et al., 1994b), plants possess chloroplastic and cytosolic NADP-ME (EC 1.1.1.40) isoforms (Drincovich et al., 2001; Maier et al., 2011).

The first two plant species in which the complete set of NADP-MEs were analyzed are the model C_3 plants *Arabidopsis thaliana* (Gerrard Wheeler et al., 2005) and rice (Chi et al., 2004). The repertoire of NADP-MEs in both species included four isoforms, with three proteins functioning in the cytosol and one in the plastid. Studies in both species have shown differences between individual isoforms in terms of expression (Chi et al., 2004; Gerrard Wheeler et al., 2005), oligomerization state and kinetic behaviour (Gerrard Wheeler et al., 2005). Significant variations observed in the enzymatic performance were unanticipated, since the degree of identity on the amino acid level composition shared by the four isoforms was high (75–90%) (Gerrard Wheeler et al., 2008). These kinetic variations also implied that minor changes in the primary structure must be sufficient for enabling the differing kinetic behaviour of each isoform.

In maize and other plants performing the NADP-ME subtype of C_4 photosynthesis, the NADP-ME isoform localized in the BS chloroplasts is the predominant decarboxylase responsible for CO_2 release from malate (Maier et al., 2011). This isoform is referred to as the C_4, or photosynthetic, NADP-ME isoform (C_4-NADP-ME) (Maier et al., 2011). In addition, some CAM plants are also known to employ this decarboxylase in order to provide CO_2 for subsequent assimilation by RubisCO during the day time (Edwards and Andreo, 1992; Drincovich et al., 2001). However, in contrast to the C_4 species, in CAM plants the photosynthetic isoform is localized in cytosol (Winter et al., 1982). As in C_3 plants, cytosolic NADP-ME isoforms have also been identified in C_4 plants. Moreover, C_4 plants additionally possess a second chloroplastic isoform, referred to as the $nonC_4$-NADP-ME (Drincovich et al., 2001; Maier et al., 2011).

3.3.2 NADP-ME protein family in maize comprises five members

Maize is the C_4 plant with the best-studied repertoire of NADP-ME isoforms due to the fact that the ZmC$_4$ (Drincovich et al., 1991) and ZmnonC$_4$ (Maurino et al., 2001) NADP-ME isoforms were readily purified directly from the plant and subsequently characterized at the molecular level. Elucidation of the maize genome (Schnable et al., 2009) has ultimately led to identification of all the *NADP-ME* genes. Apart from the two chloroplastic isoforms, three cytosolic isoforms (Zmcyt1 and Zmcyt2; (Alvarez et al., 2013) and (Zmcyt3) (Detarsio et al., 2008) were identified and characterized. The three maize genes encoding cytosolic NADP-ME isoforms display different patterns of organ expression, response to stress stimuli and the encoded proteins have different biochemical properties (Alvarez et al., 2013).

Individual isoforms of the maize NADP-ME protein family share a high degree of amino acid identity, with the lowest value of 63% shared between Zmcyt1-NADP-ME and ZmC$_4$-NADP-ME and the highest of 84% shared between the chloroplastic isoforms, suggesting that the more recent divergence occurred for the genes encoding the plastidic isoforms (Alvarez et al., 2013). The two plastidic maize NADP-MEs exhibit differences in terms of their expression, structural organization and kinetic behaviour (Alvarez et al., 2013; Saigo et al., 2013). Expression analyses performed for *ZmnonC$_4$-NADP-ME* showed its presence in all the organs analyzed with the highest accumulation in the non-photosynthetic organs (Alvarez et al., 2013). In contrast to this, *ZmC$_4$-NADP-ME* was almost exclusively present in the photosynthetic organs (Alvarez et al., 2013). In terms of the occurrence in different regions of the leaf blade, the expression of *ZmC$_4$-NADP-ME* increased from the base to the tip and thus correlated with C_4 cycle activity (Pick et al., 2011; Alvarez et al., 2013). The photosynthetic isoform is localized within BS chloroplasts (Maurino et al., 1997; Chang et al., 2012; Tausta et al., 2014). The photosynthetic and non-photosynthetic isoforms have significant differences in terms of biochemical and kinetic properties (Table 1). The photosynthetic isoforms has two-fold higher affinity for malate and nine-fold higher affinity for NADP, and two-fold higher catalytic rate than the non-photosynthetic isoform (Saigo et al., 2004; Alvarez et al., 2013). ZmC$_4$-NADP-ME is assembled as a homotetramer, whereas its non-photosynthetic counterpart exists as a homodimer (Saigo et al., 2004; Saigo et al., 2013) (Table 1). The photosynthetic isoform possesses a unique regulatory property: its activity is inhibited by higher malate concentrations in a pH-dependent manner (Detarsio et al., 2007). Recently, it was shown that the photosynthetic isoform is redox-modulated *in vitro* (Alvarez et al., 2012). This opens the possibility that post-translational modifications might be involved in the regulation of this isoform.

Table 1: Comparative summary of the most significant biochemical, structural and regulatory characteristics of the maize cloroplastic NADP-ME isoforms. Kinetic parameters taken from Alvarez et al., 2013; regulatory characteristics taken from Saigo et al., 2013.

	ZmC_4	$ZmnonC_4$
Gene identifier	GRMZM2G085019	GRMZM2G122479
k_{cat} (s^{-1})	201.3 ± 9.8	105.6 ± 8.9
K_m malate (mM)	0.22 ± 0.02	0.42 ± 0.03
K_m NADP (μM)	8.0 ± 0.3	70.2 ± 0.3
Monomer molecular mass (kDa)	62	64
Intracellular localization	Plastidic	Plastidic
Oligomerization state	Homotetramer	Homodimer
Inhibition by higher malate concentrations at pH 7.0	+	-
Redox modulation *in vitro*	+	-

3.4 Post-translational modifications as a regulatory mechanism of protein function

Post-translation modifications (PTMs) represent one of the molecular mechanisms enabling the metabolic adaptation and eventual survival of an organism subjected to changes in its environmental conditions (Mazzucotelli et al., 2008). In general, a PTM can be defined as any chemical modification of a protein as a result of a spontaneous non-enzymatical or enzymatical covalent attachment of a specific group to a specific amino acid residue(s) or termini of a protein (Seet et al., 2006; Sims and Reinberg, 2008). PTMs can be triggered by changing cellular environmental factors, such as pH changes, reactive oxygen species presence, etc. (Buchanan and Balmer, 2005). Whereas most of the enzymatically mediated PTMs (e.g. the most widely studied phosphorylation and acetylation) are reversible (Seet et al., 2006), a significant number of non-enzymatic PTMs (e.g. glycation and carbonylation) are irreversible (Harmel and Fiedler, 2018).

In general, all PTMs may change the physicochemical properties of target proteins and lead to changes in their structural organization. The last, in turn, could influence protein's turnover, cellular localization, activity or binding partners (Beltrao et al., 2013).

As of November 2018 more than 600 different PTMs were listed in the Uniprot database (Bateman et al., 2017), and were identified on proteins from organisms belonging to all domains of life (http://www.uniprot.org/docs/ptmlist). The broad range of PTMs based on the

covalent modifications of target proteins includes those functioning via attachment of the functional groups with low molecular weight, as in case of phosphorylation and acetylation, together with the PTMs based on addition of whole proteins, as in case of ubiquitylation and sumoylation (Beltrao et al., 2013; Spoel, 2018). Through this highly variable nature of PTMs, with each having distinct chemical and structural characteristics, a multiplicity of different species resulting from the same target protein can be obtained (Jungblut et al., 2008), which contributes significantly to the diversity of the proteome of an organism and increases it by orders of magnitude (Walsh et al., 2005; Venne et al., 2014).

3.4.1 Protein phosphorylation belongs to the functionally best-studied PTMs

Protein phosphorylation is defined as a reversible esterification of a high-energy gamma phosphoryl group, mainly from ATP, to a specific amino acid in the protein amino acid sequence (Day et al., 2016). In some cases, guanosine triphosphate and PEP molecules can also be utilized as phosphate donors (Hunter, 2012). Covalent phosphate attachment to the side chain of an amino acid is executed by a member of the protein kinase family of enzymes (Roskoski, 2015; Smoly et al., 2017). Removal of the phosphoryl moiety is mediated by members of another family of enzymes, protein phosphatases (Moorhead et al., 2009). A representation of this reversible phosphorylation-dephosphorylation reaction is provided in Figure 2.

In eukaryotic cells, phosphorylation mainly affects three hydroxyl-group-containing amino acids tyrosine (Tyr), threonine (Thr) and serine (Ser). Collectively, phosphorylation events at these amino acids are referred to as the O-phosphorylation sites. From these three amino acids, serine and threonine residues represent the preferred targets of protein kinases. In contrast to this, significantly smaller amounts of the total esterified phosphate are found to be attached to the tyrosine residues within eukaryotic proteins (Sharma et al., 2014; Day et al., 2016; Li et al., 2016). As an example, in the unfractionated whole cell lysates of Arabidopsis, in which 2172 phosphorylation sites from 1346 proteins were identified, the distribution of pSer, pThr and pTyr was 85.0, 10.7 and 4.3% (Sugiyama et al., 2008).

Besides phosphorylation of hydroxyl-group-harboring amino acids, phosphorylation events at six more amino acids – arginine, histidine, lysine, aspartate, glutamate and cysteine – are chemically possible and have been experimentally detected in each case except glutamic acid (Cieśla et al., 2011; Hunter, 2012). The physiological significance of phosphorylation at these residues remains less investigated in both eukaryotes and prokaryotes (Cain et al., 2014).

Historically, the first report of enzymatic protein phosphorylation as a regulatory physiological mechanism was documented in 1955 by Fisher and Krebs (Fischer and Krebs, 1955). Since then an immense variety of phosphorylation sites have been identified, and many of these sites have been characterized in eukaryotes (Olsen et al., 2006; Bi et al., 2011) and prokaryotes (Deutscher and Saier, 2005; Mijakovic and Macek, 2012).
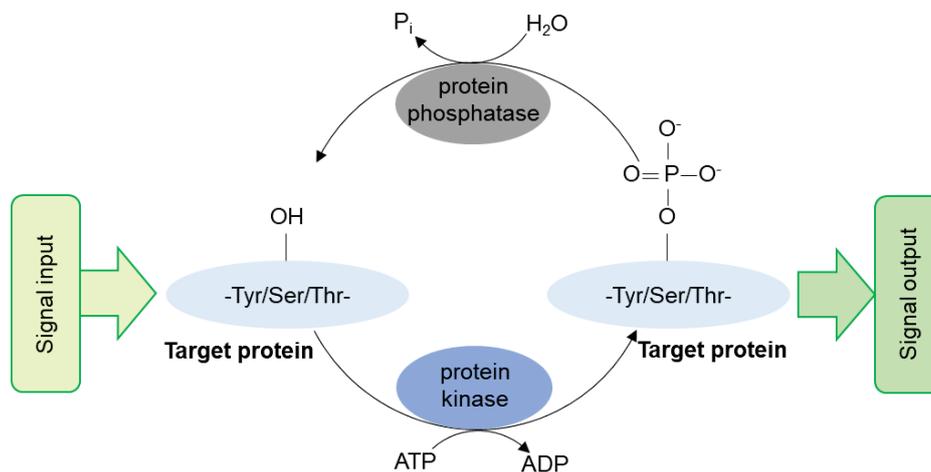
Figure 2: Reversible protein phosphorylation belongs to the enzymatically mediated PTMs and functions as a molecular switch mechanism in response to changing internal and external stimuli. Reversible protein phosphorylation is carried out via addition of a phosphoryl group (typically, from an ATP molecule) to specific amino acid residues (most commonly to serine, threonine and tyrosine), which is catalyzed by a member of a protein kinases enzyme family. For the removal of the phosphoryl moiety from target proteins, members of the protein phosphatase family are responsible. Figure modified after Humphrey et al. (2015).

Protein phosphorylation is the most extensively studied type of PTM in terms of functionality. It is of interest due to its participation in the broadest range of cellular processes and especially by its crucial role in response mediation to a plethora of different stimuli by signal transduction in pro- and eukaryotes and archaea (Zhulin et al., 2003; Humphrey et al., 2015; Xu and Zhang, 2015). Phosphorylation events were reported in the case of numerous enzymes participating in metabolic reactions (Oliveira et al., 2012), mediating cell growth (Bengoechea-Alonso and Ericsson, 2016), cell division (Fry et al., 2012), cell differentiation (Wang et al., 2016), immunity (Brodin et al., 2015; Saleh et al., 2015), circadian control (Kusakina and Dodd, 2012; Robles et al., 2017), bacterial virulence and antibiotic resistance (Sun et al., 2012), as well as many other cellular processes.

In line with the paramount importance of phosphorylation in regulating cellular functions, about one third of total eukaryotic proteins in the cell are considered to be in a phosphorylated state at any time (Zolnierowicz and Bollen, 2000). Practical results of the global phosphoproteomic studies often showed that an even higher portion of cellular proteins is

subjected to this PTM. As an example, in the human cancer cell line HeLa S3 more than three-quarters of the identified proteins were phosphoproteins (Sharma et al., 2014).

3.4.2 Phosphoproteomics is the most widely used tool for identification of protein phosphorylation

Incorporation of radioactive inorganic phosphorus atoms into cellular proteins is one of the first experimental methods to identify O-phosphorylation sites (Salih, 2005). However, this approach possesses significant drawbacks since significant amounts of radioactivity and considerable amounts of purified phosphorylated proteins are needed.

In order to overcome these limitations, new protein analytical methods based on mass spectrometry (MS) were developed (Steen and Mann, 2004; Domon and Aebersold, 2006; Richards et al., 2015). In MS experiments, biomolecules are subjected to ionization and the mass of the obtained fragments is determined by observing their specific trajectories in a vacuum system (Aebersold and Mann, 2003; Steen and Mann, 2004; Han et al., 2008). Phosphoproteomics is a based on liquid chromatography (LC)–MS branch of proteomics that aims to identify, map at single amino acid resolution, systematize, and, in some cases, quantify protein phosphorylation events in a given sample (Silva-Sanchez et al., 2015). Since phosphorylation of an amino acid residue introduces a mass shift of 80 Da relative to the unmodified peptide or protein, it can be detected in a MS-based phosphoproteomic experiment (Jensen, 2006; Witze et al., 2007).

The most widely used method for identification of phosphorylated proteins is known as the shotgun approach and belongs to the bottom-up proteomic strategies (Lothrop et al., 2013; Schwämmle et al., 2015). Identification of the amino acid residue(s) subjected to phosphorylation is the first step towards understanding how this regulatory event(s) is involved in the global cellular regulation interplay. Even though standard shotgun MS-based phosphoproteomic approaches deliver reproducible and reliable information about distinct phosphorylated residues, these techniques still cannot provide information on the stoichiometry or relative changes in phosphorylation of the given phosphosite(s), which complicates the biological interpretation of the results (Nita-Lazar et al., 2008; Schulze and Usadel, 2010; Dephoure et al., 2013).

Therefore, it is important to use additional experimental techniques for both validation of the phosphorylation site(s) discovered in large-scale phosphoproteomic studies, and for deeper understanding of their biological function (Silva-Sanchez et al., 2015). In general, significance of the newly discovered phosphorylation site(s) can be analysed on three levels – biochemical,

structural and physiological (Jünger and Aebersold, 2014). The main goal of the analysis on the biochemical level is the verification that phosphorylation, which was observed at the peptide level in MS-based phosphoproteomic studies, occurs at the protein of interest *in vivo*. A mechanistic or structural analysis represents a second level of significance of the phosphosite (Jünger and Aebersold, 2014). In this case, the main question is whether and how the observed phosphorylation event influences the activity and structure of the protein of interest and its possible interactions with different cellular proteins. This aspect remains quite challenging, since phosphoamino acids are not encoded in the genetic code. After confirmation of the occurrence of a specific phosphorylation event on a protein level and determination of its influence on the protein function, the biological significance of the phosphorylation event should be addressed (Zhu and Li, 2013; Jünger and Aebersold, 2014).

3.4.3 Phosphorylation may change the physicochemical and structural properties of target proteins and consequently alter their functions

Phosphorylation of a protein has often been shown to alter its molecular functions. This can be explained by the physicochemical and structural properties of the phosphoryl group. In particular, it has a large hydrated shell and its negative charge is always greater than 1, with the highest charge being -2 and typically being exhibited at physiological pH (Johnson and Lewis, 2001; Johnson, 2009; Hunter, 2012). These characteristics make phosphorylated amino acids chemically quite distinct from the only two genetically encoded amino acids with a strict negative charge – Asp and glutamic acid (Glu). Additionally, Asp and Glu have a smaller hydrated shell and a negative charge never exceeding -1.

In terms of interactions with molecular surrounding, a phosphoryl group attached to a protein can participate in the formation of intra- and intermolecular hydrogen bonds and salt bridges over the oxygen atoms in this functional group (Mandell et al., 2007; Nishi et al., 2014). In addition, due to the higher density of negative charges and larger size of the hydrated shell, these bonds made by phosphorylated amino acids are more stable than those made by Glu and Asp (Hunter, 2012). All of these differences transform the phosphoamino acids into new chemical units and thus give them the potential to alter the energy landscape of a protein and contribute to a diversification of its surfaces (Hunter, 2012; Xin and Radivojac, 2012).

Covalent attachment of the phosphoryl group can introduce local structural changes within a peptide motif and/or lead to global perturbation in the overall protein conformation (Chen and Cole, 2015). These structural alterations may spread to the more distantly located portions of protein and thus provoke changes in enzymatic activity, even if the PTM is introduced far from

the catalytic center; this phenomenon is termed "allosteric PTM" (Nussinov et al., 2012). Phosphorylation at the Ser14 within dimeric glycogen phosphorylase represents a prominent example of such global changes introduced to protein tertiary and quaternary structure by a single phosphosite: this residue has been shown to be shifted about 50 Å after phosphorylation, resulting in changes to contacts between subunits of the phosphorylase (Barford et al., 1991; Johnson and Lewis, 2001). In general, perturbations introduced by phosphorylation event(s) may potentially lead to modulations in the following protein functions: enzymatic activity and affinity to substrate; protein-protein interaction or binding events; stability; turnover rate and cellular localization (Johnson, 2009; Aebersold and Mann, 2016) (Figure 3).
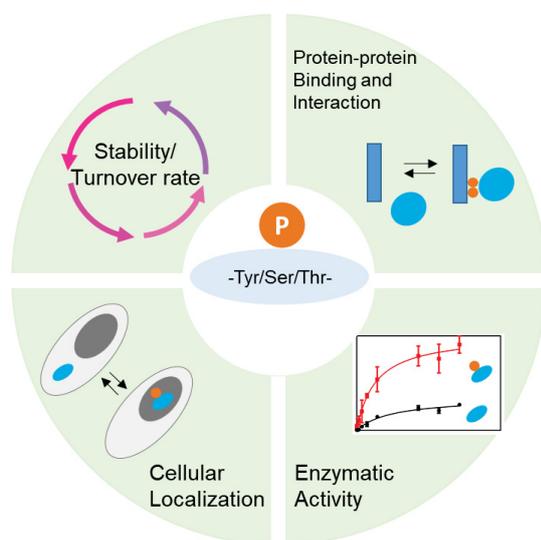
Figure 3: Schematic representation of the molecular protein functions which are affected by phosphorylation. In general, covalent attachment of the phosphoryl group to a specific amino acid residue may potentially lead to changes in protein-protein binding events and interactions, influence enzymatic activity, cellular localization and/or turnover rate of the affected protein. Figure modified after Aebersold and Mann (2016).

Phosphorylation can promote enzymatic activity, as observed for human cyclin-dependent protein kinase family member Cell Division Cycle 2 with phosphorylation at the Thr161 residue (Fesquet et al., 1993; Nigg, 1995). Phosphorylation can also decrease enzymatic activity, as in case of the isocitrate dehydrogenase from *Escherichia coli*; this enzyme was shown to be completely inactivated by phosphorylation at Ser113 (Thorsness and Koshland, 1987). Interestingly, comparison of the crystal structures from the non-phosphorylated and phosphorylated forms of this enzyme has shown that phosphorylation at this residue does not lead to large conformational changes but rather inhibited enzymatic activity by a direct

electrostatic interaction between the substrate isocitrate and the phosphate moiety attached to the serine residue (Hurleys, 1990).

Phosphorylation at a specific site(s) has also been shown to "label" affected proteins to subsequent proteolysis for ubiquitin-dependent protein degradation. An example is the transcription coactivator nonexpressor of pathogenesis-related genes 1 (NPR1) (Spoel et al., 2009). NPR1 is known as a master regulator of plant immunity and its nuclear translocation is a critical regulatory step in development of systemic acquired resistance (SAR). For prevention of the untimely activation of SAR, NPR1 is continuously removed from the nucleus by proteasome, and its coactivator activity is restricted. Inducers of SAR promote NPR1 phosphorylation at residues Ser11 and Ser15, after which a Cullin3-based ligase rapidly ubiquitinylates NPR1, which targets this protein for degradation by the proteasome (Spoel et al., 2009).

Phosphorylation can also play an important role in governing the subcellular localization of a protein. Lee and coauthors showed that Arabidopsis NPR1 is phosphorylated at Ser589 and Thr373 by SNF1-related protein kinase 2.8 during SAR and that these phosphorylation events are important for its nuclear import (Lee et al., 2015).

Phosphorylation also plays a crucial role in the modulation of the nature and strength of protein-protein interactions (Johnson, 2009; Jin and Pawson, 2012). Phosphoryl groups covalently attached to some amino acids can be recognized by specific phospho-Ser/Thr or Tyr binding domains. An example of such domains is provided by the Src-homology 2 domain which recognizes and binds to the specific phosphotyrosine sites (Pawson et al., 2001).

These examples show that physicochemical and structural consequences of phosphorylation event(s) are diverse and that their effect(s) on protein characteristics are not easily predictable.

3.4.4 Regulatory roles of phosphorylation are known for a few C$_4$ proteins

In terms of the dual-cell based C$_4$ cycle in general and maize in particular, PPDK, PEPC and PEP-CK were reported to be regulated by phosphorylation and the exact functions of these phosphorylation events have been elucidated (Wang et al., 2011).

Maize PEPC was shown to undergo reversible regulatory phosphorylation at the conserved Ser15 residue near the N-terminus of the protein under illumination (Jiao and Chollet, 1988; Jiao and Chollet, 1989; Jiao and Chollet, 1990). This phosphorylation enhances the catalytic activity and concomitantly decreases sensitivity to the allosteric inhibitor malate (Jiao and Chollet, 1988). PEPC phosphorylation starts before dawn, reaches a maximum by mid-morning and starts to decrease again before dawn (Ueno et al., 2000). In addition to maize PEPC, its

homolog from sorghum was also shown to be regulated by phosphorylation at Ser8 (Bakrim et al., 1992), and PEPC from CAM plants was also reported to be regulated by phosphorylation (Nimmo et al., 1990).

PPDK activity is known to be regulated by phosphorylation of an active-site threonine residue (Thr527); phosphorylation of this residue leads to a decrease in enzymatic activity (Ashton and Hatch, 1983; Chen et al., 2014). Light activity rather than light/dark transition or circadian rhythm was shown to be responsible for regulation of phosphorylation at this residue (Chen et al., 2014). Both phosphorylation and dephosphorylation of this residue are catalyzed by a unique bifunctional enzyme PPDK regulatory protein (Burnell and Hatch, 1985; Burnell and Hatch, 1986), which shares no significant sequence similarity with other protein kinases or phosphatases (Burnell and Chastain, 2006; Chastain et al., 2008). The structural mechanism of reversible PPDK phosphorylation by this enzyme has recently been deciphered (Jiang et al., 2016).

Phosphorylation of PEP-CK was first described in the C$_4$ species *Panicum maximum* (Guinea grass), in which it was phosphorylated in the darkened and dephosphorylated in illuminated leaves (Walker and Leegood, 1996). The non-phosphorylated PEP-CK exhibited a higher affinity for its substrates OAA and PEP (Walker et al., 2002). Additionally, using phosphoproteomics and sequence-specific antibodies, maize PEP-CK was shown to be phosphorylated at Ser55, Thr58, Thr59 and Thr120 (Chao et al., 2014). The amount of Ser55 affected by phosphorylation dramatically increased upon illumination of maize seedlings and decreased in the leaves of seedlings transferred from light to darkness. Ser55 phosphorylation was also shown to affect enzymatic performance of maize PEP-CK, since the decarboxylase activity of PEP-CK in illuminated leaves was lower when compared to the activity in the darkened leaves (Chao et al., 2014).

This phosphorylation event might represent a potential off-switch for PEP-CK decarboxylase activity and, by this, participate in the coordination of activity of two decarboxylases in maize BS cells: phosphorylation at Ser55 under illumination, when the most of the CO$_2$ is derived from malate decarboxylation by NADP-ME, leads to decrease in the PEP-CK activity in maize BS cells. In the dark, when the activity of the NADP-ME decarboxylation pathway is lower, dephosphorylation of PEP-CK provokes an increase of its enzymatic activity, which leads to an enhanced carbon dioxide production via the aspartate decarboxylation pathway (Chao et al., 2014).

Considering the rapid advances in the MS-based phosphoproteomic techniques, it is inevitable that a wider number of implicated in C$_4$ photosynthesis regulation PTMs, and phosphorylation events in particular, will be identified.

4 Aims of this thesis

The constantly growing gap in the yields of primary crops, which is, among other things, provoked by global climate changes (Wheeler and von Braun, 2013), is an important challenge for modern society (Conceição and Mendoza, 2009; Ray et al., 2013). Engineering certain traits or the complete C_4 pathway into C_3 crops represents one of the most promising strategies for solving this problem (von Caemmerer et al., 2012; Long et al., 2015). At present, fundamental biochemical and anatomic features of the C_4 pathway are well understood, but many evolutionary and regulatory mechanisms of this syndrome remain elusive. Knowledge about these aspects is relevant for creating a truly efficient artificial C_4 crop (von Caemmerer et al., 2012). One promising approach would be introducing the two-celled C_4 system based on the NADP-ME biochemistry into C_3 crops. This subtype of photosynthesis is utilized by many of the most productive C_4 crops and is relatively simple (Covshoff and Hibberd, 2012).

Accordingly, the first aim of this thesis is the identification of the signature amino acid substitutions in the C_4-NADP-ME from grasses exhibiting the NADP-ME subtype of C_4 photosynthesis, and maize in particular. This is followed by an in-depth analysis of the roles of the identified positions in conferring increased affinity to malate, a pH-dependent inhibition by malate, and homotetrameric assembly – all unique C_4 properties of this decarboxylase.

The second aim of this thesis is the investigation of post-translational regulations of the maize C_4-NADP-ME using discovery by MS and assessment of the biochemical, structural and physiological importance of the identified PTM(s).

5 Manuscripts associated with this thesis

5.1 Manuscript I: Molecular adaptations of NADP-malic enzyme for its function in C$_4$ photosynthesis in grasses

Authors: **Anastasiia Bovdilova***, Astrid Höppner*, Clarisa E. Alvarez*, Christian-Claus Wolff, Mariana Saigo, Tao Zhang, Luitgard Nagel-Steger, Maria F. Drincovich, Martin J. Lercher, Veronica G. Maurino

* These authors contributed equally to the work

Contribution of Anastasiia Bovdilova:

In vitro mutagenesis and cloning of expression constructs

Production of recombinant enzymes

Measurement of kinetic data

Evaluation of circular dichroism measurements

Production of the ZmC$_4$-NADP-ME crystals

Figure design and drafting of the manuscript

Molecular adaptations of NADP-malic enzyme for its function in C$_4$ photosynthesis in grasses

Anastasiia Bovdilova[1,2§], Astrid Höppner[2,3§], Clarisa E. Alvarez[4§], Christian-Claus Wolff[1,2], Mariana Saigo[4], Tao Zhang[6], Luitgard Nagel-Steger[6], Maria F. Drincovich[4], Martin J. Lercher[2,7], Veronica G. Maurino[1,2,*]

[1]Plant Molecular Physiology and Biotechnology Group, Institute of Developmental and Molecular Biology of Plants, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[2]Cluster of Excellence on Plant Sciences (CEPLAS), 40225 Düsseldorf, Germany.

[3]Center for Structural Studies (CSS), Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[4]Centro de Estudios Fotosinteticos y Bioquimicos (CEFOBI-CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, University of Rosario, Suipacha 570, 2000 Rosario, Argentina.

[5]Laboratory of Molecular and Structural Microbiolgy, Institute Pasteur Montevideo, Mataojo 2020, 11400 Montevideo, Uruguay.

[6]Institut für Physikalische Biologie, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany and Institute of Complex Systems, Structural Biochemistry (ICS-6), Forschungszentrum Jülich, 52425 Jülich, Germany.

[7]Institute for Computer Science and Department of Biology, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[§]These authors contribute equally to the work

[*]Corresponding author: Veronica G. Maurino (veronica.maurino@uni-duesseldorf.de). Plant Molecular Physiology and Biotechnology Group, Institute of Developmental and Molecular Biology of Plants, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany. Tel.: 49-211-8112368.

ABSTRACT

In C_4 grasses of agronomical interest, malate shuttled into the bundle sheath cells is mainly decarboxylated by an isoform of NADP-malic enzyme (C_4-NADP-ME). The activity of C_4-NADP-ME was optimized in evolution to efficiently deliver CO_2 to Rubisco. C_4-NADP-ME evolved through the duplication of a closely related plastidic non-photosynthetic NADP-ME (nonC_4-NADP-ME). During its evolution, C_4-NADP-ME acquired distinct kinetic and biochemical features, such as increased efficiency, tetrameric structure, and pH-dependent inhibition by its substrate malate. Here, we identified the molecular basis of those C_4 adaptations based on strict differential conservation of amino acids between nonC_4- and C_4-isoforms sequences, combined with solving the first plant crystal structures of maize C_4-NADP-ME as apoenzyme and sorghum C_4-NADP-ME in complex with its cofactor NADP. Site-directed mutagenesis of the four identified amino acids of maize C_4-NADP-ME to those conserved in the non-photosynthetic isoforms confirmed their involvement in conferring crucial C_4 properties. Phe140 is critical for the acquisition of the active tetrameric structure, while Gln503, Leu544, and Glu339 independently confer high affinity for malate; in addition, Glu339 is sufficient to confer high catalytic efficiency as well as the allosteric regulation by malate. Together, these four adaptive substitutions formed the basis for the efficient catalysis and regulation of one of the central biochemical steps in C_4 metabolism. The small number of mutations required to optimize NADP-ME for the C_4 function may well have contributed to the repeated emergence of NADP-ME C_4 subtype.

C_4 photosynthesis is a complex trait that confers strongly reduced photorespiration and greater photosynthetic efficiency, accompanied by increased nitrogen and water use efficiency, in plants growing in temperate or hot climates and under high irradiation [1-3]. Despite its biochemical and anatomical complexity, C_4 photosynthesis emerged independently more than 60 times; it is particularly prevalent in grasses, where it evolved in approx. 18 lineages [4]. The evolution of the C_4 trait involved an increase in the number and size of bundle sheath cells (BSC), a higher organelle content of BSC, and the cellular re-location and molecular adaptation of photosynthetic and photorespiratory enzymes [3]. The adaptation of cellular activities to the C_4 syndrome regularly involved gene duplications and neo-functionalization, as well as changes in expression and biochemical properties of existing gene products; however, many details of these crucial evolutionary changes are still obscure [5].

The largest C_4 biochemical subtype, which includes the major monocot crops, *Zea mays* (maize) and *Sorghum bicolor* (sorghum), uses a C_4-specific isoform of NADP-dependent malic enzyme (C_4-NADP-ME) to decarboxylate malate in the BSC chloroplasts [6]. Optimized kinetics and tight regulation of this enzyme are essential to C_4 photosynthesis as coordination of C_4-NADP-ME decarboxylation and Rubisco carboxylation rates is imperative to maintain an effective flux through the C_4 pathway. The C_4-NADP-ME evolved through the duplication of a gene encoding a plastidic non-photosynthetic isoform (nonC_4-NADP-ME), followed by the acquisition of a BSC-specific expression pattern [7-10]. In monocots, this gene duplication event occurred prior to the divergence of the Poaceae [5]. This means that the same duplicated NADP-ME was co-opted for the C_4 function in independent C_4 lineages [5].

Plastidic C_4- and nonC_4-NADP-ME isoforms differ in their kinetic and allosteric properties. In the monocot grasses maize and sorghum, C_4-NADP-ME has catalytic efficiencies that are 2 to 5 times higher than those of the nonC_4-isoforms [11]. The affinity of C_4-NADP-ME for malate is twofold higher than that of the nonC_4-isoform (K_m malate: 0.23 mM vs. 0.42 mM in maize [8], 0.36 mM vs. 0.62 mM in sorghum [11]). In both species, the C_4-NADP-ME isoform is inhibited by malate at pH 7.0 [8,11-14]. The affinity of C_4-NADP-ME for the cofactor NADP is 2 to 10 times higher than that of the nonC_4-isoforms [8,14].

The C_4- and nonC_4-NADP-ME of maize and sorghum also differ in their quaternary organization: the photosynthetic isoform is assembled as a tetramer, while the non-C_4 isoform assembles as a dimer [7,8,11,15]. The particular kinetic properties and oligomeric organization of C_4-NADP-ME seem to be crucial features for C_4 optimization of this decarboxylase, as they

were also described for the independently evolved C_4-NADP-ME isoforms of the monocot grass *Saccharum officinarum* (sugar cane) [13] and the eudicot *Flaveria sp.* [12].

Why are the allosteric regulation of C_4-NADP-ME by malate and the formation of a tetramer C_4 adaptations? During the night, when the stromal pH decreases to around 7.0 - 7.5 [16], C_4-NADP-ME can be inhibited by its own substrate, a property that confers this isoform a regulation of its activity when photosynthesis is not active. Thus, malate inhibition of C_4-NADP-ME activity in darkness is key for preventing an extreme decrease in cellular malate content. This ensures a rapid kick-start for Rubisco action at the beginning of the next light period and avoids carbon starvation during the night period [17,18]. In addition to the allosteric regulation, a proportion of C_4-NADP-ME likely loses its quaternary structure during the night and adopts a less active or inactive lower oligomerization state [13]. Thus, the tetrameric organization of C_4-NADP-ME forms the basis for a second level of enzymatic regulation. However, the molecular details of C_4-NADP-ME regulation are still unclear.

The structural and biochemical changes that occurred during the adaptation of C_4-NADP-ME to its central role in C_4 photosynthesis must have been acquired through specific amino acid substitutions. To identify these substitutions in maize and sorghum C_4-NADP-ME, we developed an algorithm for the detection of amino acid residues that are strictly differentially conserved in plastidic photosynthetic and non-photosynthetic isoforms within the Panicoideae clade of the Poaceae family. This analysis, aided by the determination of the crystal structures of maize and sorghum C_4-NADP-ME, allowed us to identify four amino acids that confer C_4 properties on NADP-ME, such as its tetrameric organization, high affinity to the substrate, and allosteric regulation (malate inhibition).

RESULTS

Determination of the crystal structure of maize and sorghum C_4-NADP-ME

To aid identifying amino acid residues that contribute to the C_4 optimization of NADP-ME, we solved the crystal structures of maize and sorghum C_4-NADP-ME (ZmC$_4$-NADP-ME, SbC$_4$-NADP-ME). The maize enzyme structure was determined at 2.2 Å resolution, that of sorghum at 2.0 Å. Data collection and refinement statistics are summarized in Supplementary Table 1.

The quaternary structure of the ZmC$_4$-NADP-ME displays a homo-tetrameric assembly composed of two dimers (monomers A and B and monomers C and D, respectively; Fig. 1a and b). For an in-depth description of the overall structure and tetrameric assembly refer to

Supplementary information and Supplementary Fig. 1. The dimers are tilted relative to each other at an angle of ~55°. Residues forming the tetramer interface are mainly located in subunits A and C (Supplementary Table 2). The SbC$_4$-NADP-ME is likewise assembled as a tetramer (Fig. 1c), but only two monomers (A and B) are traceable unambiguously in the electron density for the entire monomer (Supplementary information). We therefore use chains A and B for further structural comparisons. The overall structure of SbC$_4$-NADP-ME is highly similar to that of ZmC$_4$-NADP-ME. The cofactor NADP is seen to be bound to both monomers A and B in the SbC$_4$-NADP-ME structure (Fig. 1c). Amino acids involved in cofactor binding and hydrogen bonding distances are given in Supplementary Fig. 2 and Supplementary Table 3.

In ZmC$_4$-NADP-ME, monomers A and D show nearly identical conformations (RMSD 0.164 Å over 496 Cα atoms), likewise monomers B and C (RMSD 0.300 over 506 Cα atoms), whereas the two monomers of one dimer reveal few but distinct differences in their conformations (A vs. B: RMSD 1.332 Å over 496 Cα atoms; C vs D: RMSD 1.145 Å over 495 Cα atoms). A superposition of ZmC$_4$-NADP-ME monomer A and B indicates that the secondary structure elements from domains A and B adopt mainly the same relative spatial position, whereas domain C in monomer B is shifted outwards (Supplementary Fig. 3a; for details see Supplementary information). We therefore term monomer A and D being in a "closed" conformation, and monomer B and C in an "open" conformation. In SbC$_4$-NADP-ME, we found a very similar arrangement: monomer A adopts an "open", monomer B a "closed" conformation (RMSD monomer A and B: 0.988 Å over 509 Cα atoms) (Supplementary Fig. 3b). The presence of these two conformations in both the apo-enzyme (ZmC$_4$-NADP-ME) and the secondary complex with NADP (SbC$_4$-NADP-ME) is surprising, as one would rather expect such a conformational change upon substrate and/or cofactor binding. From our results, the "open" and "closed" conformations seem to appear independent from ligand binding, which might hint at an alternative mode of activity.

Differential conservation as a means to identify amino acid residues involved in C$_4$ optimization of C$_4$-NADP-ME

To identify amino acid residues that contribute to the observed C$_4$-specific function of C$_4$-NADP-ME isoforms, we searched for positions in the protein sequence alignment at which all C$_4$ isoforms share the same amino acid, while all non-photosynthetic isoforms from C$_4$ (nonC$_4$) or C$_3$ plants share a different amino acid. Such a pattern of strict differential conservation would

indicate that the respective amino acid position is functionally important in all isoforms, while the selected amino acid - and hence the molecular function - differs between C_4 and nonC$_4$/C$_3$ isoforms.

We first focused on the single C_4 lineage Andropogoneae within the Panicoideae subfamily of the Poaceae, which includes the C_4 grasses maize and sorghum. As a C_3 reference species, we also included *Oryza sativa* (rice), another member of the Poaceae. Based on an alignment of plastidial C_4- and nonC$_4$-NADP-ME sequences, we identified 20 amino acids as being differentially conserved between the analysed isoforms (Supplementary Fig. 4).

To focus on the most promising of these 20 candidates for further functional analysis, we used the crystal structures of ZmC$_4$-NADP-ME and SbC$_4$-NADP-ME. Based on analysis of the structures, we identified the subset of sites that (i) could affect substrate/cofactor binding or catalysis, (ii) might be involved in dimer or tetramer formation, or (iii) might be part of a putative allosteric malate inhibition site. These candidates were F140, N142, N159, T163, R201, E339, P433, Q503 and L544 (Fig. 2).

To further restrict the number of differentially conserved amino acids identified in the Andropogoneae lineage to the most promising candidates, we extended our analysis to *Setaria italica*, another member of the Panicoideae subfamily of the Poaceae that belongs to an independent C_4 linage in the Paniceae [10,19]. We reasoned that *Setaria's* unique plastidic NADP-ME sequence (SiC$_4$-NADP-ME) has also been optimized for its role in C_4 photosynthesis. We found that only four of the amino acid positions of putative functional importance are strictly differentially conserved in the extended alignment (F140, E339, Q503, and L544 in ZmC$_4$-NADP-ME; Fig. 2). All these amino acids identified in the C_4-NADP-ME were originated through single point mutations of the respective codons in the nonC$_4$-isofom as follows: the DNA triplets coding for Glu (GAA, GAG) can be obtained from those encoding Ala (GCA, GCG), those coding for Phe (UUC, UUU) can be obtained from those encoding Ile (AUC, AUU), those coding for Gln (CAA, CAG) can be obtained from those encoding Glu and those coding for Leu (UUA, UUG) can be obtained from those encoding Phe by single point mutations. We thus decided to focus on these four sites for mutagenesis and further biochemical and structural characterization.

We produced mutant variants of ZmC$_4$-NADP-ME, in which we independently substituted the four amino acids for those conserved in the nonC$_4$-NADP-MEs: F140I, E339A, Q503E, and L544F. The recombinantly expressed and purified proteins showed the expected molecular

masses (66 kDa; Fig. 3a and Supplementary Fig. 5). All recombinant enzymes presented similar dependencies of the activity on the pH of the reaction media, with a pH optimum between 7.5 and 8.0 (Fig. 3b).

The determination of the turnover number (k_{cat}) indicated that the E339A, Q503E, and L544F mutants behave like the wild-type (WT) enzyme in terms of their catalytic rates (k_{cat} = 28.1 ± 2.3), while F140I presented an 8-fold reduced value (k_{cat} = 3.6 ± 1.0) (Fig. 3c to f). We observed no significant differences in the affinity to the cofactor NADP for E339A, Q503E, and L544F compared to the WT (Fig. 3c and f), suggesting that the conformation of the NADP binding sites is not altered in these mutants. We observed a higher affinity of F140I for NADP (Fig. 3c and f). The affinity to the substrate malate was decreased by approximately 2-fold in all mutants (Fig. 3d and f); the K_m values were similar to that of the maize nonC4-NADP-ME (K_m malate 0.43 mM; [8,15]).

These results confirm our hypothesis that the four amino acids identified based on their position in the protein structure and their differential conservation in C4 and non-C4 isoforms contribute to the functional adaptations of NADP-ME for its role in C4 photosynthesis.

We studied the inhibitory effect of malate on the enzymatic activity at pH 7.0 and found that the E339A substitution provoked a loss of this regulatory property (Fig. 3e and f). The E339A variant behaves thus like the non-C4 isoforms with respect to malate at pH 7.0. The variants F140I, Q503E, and L544F showed inhibition by malate at pH 7.0 with K_i values similar to that of the WT ZmC4-NADP-ME (Fig. 3e and f).

To further analyse the role of the C4 amino acid E339 in the allosteric inhibition by malate, we introduced this residue into the ZmnonC4-NADP-ME, producing the mutant A339E. The kinetic characterization of the recombinantly expressed and purified mutant at pH 8.0 indicated that the affinity for NADP and malate at pH 8.0 resembles that of the ZmC4-NADP-ME (Fig. 4a and b). At pH 7.0, the A339E variant gained malate inhibition, with a K_i value two-times higher than that of ZmC4-NADP-ME (Fig. 4a and c). These results confirm a pivotal role of the evolutionary A339E substitution in the acquisition of malate inhibition by the C4 isoform.

Oligomeric organization of ZmC4-NADP-ME variants

To analyse if the purified recombinant ZmC4-NADP-ME folded correctly and if any of the single mutations introduced in the engineered enzymes affected their secondary conformations, all proteins were analysed by circular dichroism (CD) spectroscopy. The obtained CD spectra

indicated that the engineered proteins presented only minor differences relative to the WT, and thus all proteins have largely conserved secondary structures (Supplementary Fig. 6).

Recombinant ZmC4-NADP-ME and the E339A, Q503E, and L544F variants assembled as tetramers at pH 8.0, as evaluated by native PAGE followed by immunoblot analysis using specific antibodies against ZmC4-NADP-ME (Fig. 5a). The F140I variant was found as a mixture of oligomeric states from tetramer to monomer (Fig. 5a). Native PAGE of the ZmC4-NADP-ME variants assayed for NADP-ME activity showed that the WT enzyme and all mutated variants are active only as tetramers in the assay conditions (the proteins were in a solution at pH 8.0) (Fig. 5b). This is consistent with the finding that only the tetrameric form of C4-NADP-ME is active in sorghum [20]. In line with the reduction of the tetramer fraction in the F140I mutant, we observed only a faint band showing NADP-ME activity in the *in-gel* assay compared to the other variants (Fig. 5b) and an 8-fold lower maximum catalytic activity relative to the WT (Fig. 3f). Together, these results indicate that the low k_{cat} of the F140I mutant is not due to lower kinetic performance of this enzyme but could be related to the smaller fraction of the enzyme that is present in the active tetrameric structure.

In order to further analyse the protein quaternary organization, the WT enzyme and the F140I variant were subjected to analytical ultracentrifugation (AUC). AUC measurements confirmed a stable tetrameric state of the WT ZmC4-NADP-ME at pH 8.0, with a sedimentation coefficient of approximately 10 S (Fig. 5c). The F140I variant showed a modified behaviour; two broad peaks with sedimentation coefficient of approximately 5S and 10 S are observed (Fig. 5c). It is important to mention that the fraction of F140I found in these peaks is much lower compared to the WT (Fig. 5c shows normalized values). This is due to the presence of protein agglomerates that form under the conditions of the assay at sedimentation coefficients higher than 20 S. With F140I, 81% of the total protein was always obtained as protein agglomerates, whereas WT ZmC4-NADP-ME protein agglomerates only represent 29.6% of the total protein.

We additionally examined the oligomerization state of the recombinant WT ZmC4-NADP-ME after overnight incubation at pH 7.0 before separation in native PAGE. Detection of the protein by immunoblot analysis indicated the presence of a low oligomeric form in the WT with a gel mobility similar to that observed for the F140I variant (Fig. 5d).

To further analyse the role of F140 in the stabilization of the oligomeric structure of ZmC4-NADP-ME, we introduced this amino acid into ZmnonC4-NADP-ME, producing the nonC4-

I140F mutant. The kinetic characterization of the recombinantly expressed and purified mutant indicated that the affinity for NADP and malate are lower than those of the ZmC4- and ZmnonC4-NADP-ME WT isoforms, while the maximum catalytic activity is increased (Fig. 4a). At pH 7.0, the I140F mutant is not inhibited by malate (Fig. 4c). Native PAGE assayed for activity showed that both the I140F and the A339E mutants present the same mobility as the WT ZmnonC4-NADP-ME, indicating that they assemble as active dimers (Fig. 4d).

DISCUSSION

We showed that the change at amino acid 140 from Phe to Ile in ZmC4-NADP-ME is sufficient to affect its quaternary structure (Fig. 5a). The F140I mutation is near the N-terminus and is located close to the dimer interface, but it is not directly involved in interactions with the neighboring protein chain (Fig. 2 and Suppl. Table 2). Instead, Phe140 is part of helix αA3, which is involved in monomer-monomer-interactions at the dimer interface. The neighboring amino acids Lys138 and Lys139 form hydrogen bonds with Glu288 and Gly196 of the adjacent protein chain (Fig. 6a). We therefore speculate that replacement of Phe140 to Ile disturbs the helical arrangement of that structure element (αA3), which in turn destroys the hydrogen bonding pattern of these two Lys residues, resulting in disintegration of the dimer. In the ZmC4-NADP-ME, F140 is necessary for a complete tetramerization; it may facilitate the generation of tetramers by stabilizing the dimers.

A chimera of C4-NADP-ME containing the primary sequence of the nonC4-NADP-ME between amino acid residues 102 and 247 forms exclusively dimers [15]. On the other hand, a chimera of the nonC4-NADP-ME containing the primary sequence of the C4-NADP-ME between residues 102 and 247 is found in both oligomeric states, as dimer and tetramer [15]. Thus, it appears likely that in addition to amino acid 140, other amino acid residues may be necessary to stabilize the oligomeric conformation of the C4-NADP-ME isoforms. Previous work conducted *in vitro* on sugarcane C4-NADP-ME showed that a change of pH from 8.0 to 7.0 provokes a dissociation of the tetrameric enzyme into a dimer [13]. Here, we showed that after long incubation of the ZmC4-NADP-ME at pH 7.0, part of the enzyme is also found in an inactive low oligomeric state (Fig. 5d). During the night, when the C4 pathway is not active and the stromal pH decreases to 7.0, the C4-NADP-ME loses its quaternary structure and adopts an inactive oligomerization state (dimer or monomer). These changes in oligomerization thus represent a way to modulate the enzymatic activity depending on the metabolic demand. The

use of a higher oligomeric organization for enzyme regulation was already suggested for the human mitochondrial NAD(P)-ME. This enzyme is found in an equilibrium of dimers and tetramers. Dissociation of the human mitochondrial NAD(P)-ME tetramer into dimers causes the enzyme to be less active, less cooperative for malate binding, and less responsive to its allosteric regulator [21].

The protein region that extends from amino acid residue 248 to the C-terminus of C4-NADP-ME is involved in the allosteric inhibition of malate at acidic pHs [15]. The lack of inhibition of the C4-NADP-ME at pH 8.0 and of the nonC4 isoform at both pHs suggests that the amino acid residue(s) involved in the allosteric regulation change their protonation state(s) depending on the surrounding pH. We found that Glu339 of ZmC4-NADP-ME is directly involved in the inhibition of malate at pH 7.0-7.5. In fact, the nonC4-isoforms possess a hydrophobic Ala at this position, while the C4 isoforms possess the polar and acidic residue Glu. We showed that the change from Glu to Ala in ZmC4-NADP-ME is sufficient to completely eliminate the allosteric regulation of this isoform (Fig. 3), and vice versa that the change from Ala to Glu in the nonC4-NADP-ME is sufficient to gain malate inhibition (Fig. 4). The crystal structures of Zm- and SbC4-NADP-ME show that Glu339 is far away from the catalytic site (Fig. 6d). Glu339 is located on the surface of helix αB4 with the side chain pointing outwards to the surrounding medium (Fig. 2); we conclude that this is the location of the allosteric binding site. The backbone and side chain of Glu339 are only involved in intramolecular hydrogen bonds with neighboring residues on the same helix (in all chains to Phe335; in chain A additionally to Lys343, in chain D to Asp336; Fig. 6b and c). As mentioned above, during the night the stromal pH decreases to 7.0-7.5. In this milieu Glu339 may enable malate to enter or to bind to the allosteric site under this condition but not during the day when photosynthesis is active. The lack of malate inhibition at pH 8.0 facilitates a high flux of malate decarboxylation to provide CO_2 for photosynthetic fixation; however, such flux during the night, at pH 7.0-7.5, would simply produce carbon loss. In a similar way, the activity of C4-PEPC is also inhibited by malate [22]; however, the light-dependent phosphorylation releases this inhibition rendering a more active enzyme when the photosynthetic flux is high.

The amino acid changes Q503E and L544F independently introduced into ZmC4-NADP-ME produced enzyme variants with reduced affinity for the substrate malate. The K_m values measured were around twofold higher that of the WT ZmC4-NADP-ME and thus resemble the K_m of the nonC4-isoform (K_m malate 0.43 mM; [8]). These results show that the residues identified by our sequence-based algorithm represent adaptive changes that gave rise to the

increased affinity of C4-NADP-ME for its substrate. During neo-functionalization of the chloroplastic NADP-ME, Glu503 of the nonC4-isoform changed to a Gln. This change involved the introduction of a (polar) uncharged amino acid in a position that is located at the surface of the protein (Fig. 2). From the crystal structures of Zm- and SbC4-NADP-ME, it can be deduced that Gln503 is positioned on helix αC5 on the protein surface (Fig. 2), forming intramolecular interactions with adjacent amino acids (Thr500, Trp507; in chain A additionally to Arg468, in chain C to Thr506). The other adaptive change involved the introduction of the aliphatic Leu at position 544, which was occupied by the aromatic Phe in nonC4-NADP-ME. The crystal structures of ZmC4- and SbC4-NADP-ME show that Leu544 is part of helix αB5 buried inside the protein. This residue participates via backbone interactions with Leu548 and furthermore with Val549 (in chain A), Gly547 (in chains B, C), Ile540 (in chains B, C, D), and Phe541 (in chains C, D). Helix αB5 contains two amino acids (Asn536 and Asn537) that participate in the binding of malate at the catalytic site. Our results suggest that a change in the chemical properties at amino acid position 544 induced structural changes in αB5 that influenced these enzyme-substrate interactions.

In sum, our work defines molecular adaptions that occurred during the evolution of C4-NADP-ME from a housekeeping isoform. We found that Phe140 is critical for the acquisition of the active tetrameric structure of C4-NADP-ME, while Gln503, Leu544, and Glu339 independently confer high affinity for malate; in addition, Glu339 is sufficient to confer high catalytic efficiency as well as the allosteric regulation by malate. Our results indicate that the evolution of the C4 isoform is the result of a small number of point mutations rather than of extensive sequence changes. Interestingly, each molecular adaption identified in C4-NADP-ME originated through a single point mutation of the respective codon in the nonC4-isofom. One of the major open questions in C4 evolution is why the NADP-ME subtype evolved much more frequently than the NAD-malic enzyme subtype. The small number of mutations required to optimize NADP-ME for the C4 function may well have contributed to the repeated emergence of this subtype.

To enhance plant yield by installing the C4 trait into a C3 plant [23-26], it is not only important to guarantee a high and localized expression of the C4 enzymes, but also to fine tune their activities to facilitate the required photosynthetic flux through the C4 cycle during the day while avoiding carbon loss during the night. Our results show that at least for NADP-ME, this important goal may be achieved by introducing a small number of mutations into a copy of the C3 plant's own enzyme rather than by introducing a foreign C4 isoform.

METHODS

Sequence selection and identification of C$_4$-specific amino acids

We obtained the amino acid sequences of the plastidic C$_4$- and nonC$_4$-NADP-ME isoforms of *Z. mays* cultivar Ensembl-18 (Sequence IDs: GRMZM2G085019_T01, GRMZM2G122479_P01) and *S. bicolor* cultivar v3.1.1 (Sequence IDs: Sobic.003G036200.1, Sobic.009G108700.1) from the Phytozome 11 database (Goodstein et al., 2012). We used the maize C$_4$ and nonC$_4$-NADP-ME sequences as BLAST queries to identify the corresponding orthologs in the *Setaria italica* cultivar v2.2 proteome in the same database (Sequence ID: Seita.5G134300.1). We also used the maize nonC4-NADP-ME sequence as a BLAST query to identify the unique plastidial NADP-ME in the *Oryza sativa* cultivar v7_JGI proteome (LOC_Os01g09320.1).

All sequences were aligned using PRANK (Loytynoja & Goldman 2005). We used the alignment of C$_4$- and nonC$_4$-NADP-ME isoforms (Supplementary Fig. 4) to identify amino acid positions that are strictly differentially conserved in C$_4$ compared to nonC$_4$ sequences: we selected all positions for which the C$_4$ sequences of *Z. mays* and *S. bicolor* contained the identical amino acid, while a second, different amino acid was found in the nonC$_4$ sequences of *Z. mays* and *S. bicolor* as well as in the unique plastidial *O. sativa* sequence. To identify the subset of the most promising candidates of strictly differentially conserved amino acid positions, we repeated this analysis; now, we additionally required that the corresponding amino acid of the unique plastidial NADP-ME sequence of *S. italica* was identical to that of the C$_4$-NADP-ME of *Z. mays* and *S. bicolor*.

Generation of expression constructs and site-directed mutagenesis of ZmC$_4$-NADP-ME and ZmnonC$_4$-NADP-ME

To express ZmC$_4$-NADP-ME, the cDNA sequence coding the mature protein was amplified from the plasmid pET32*::*ZmC$_4$ NADP-ME [14] and cloned in the pET16b expression vector (Novagen) using the Gibson assembly [27] and the following primer pairs: "Gib_FW" (5´-GCCATCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGT CGTCATGCGATGGTCTC-3´) and "Gib_RV" (5´-GCAGCCAACTCAGCTTC CTTTCGGGCTTTGTTAGCAGCCGCTACCGGTAGTTGCGGTAGAC-3´). The resulting expression construct, *pET16b::ZmC$_4$-NADP-ME*, was used to introduced the desired point

mutations by site-directed mutagenesis. Amplification of the whole expression construct was conducted in a PCR assay using the Physion Hot Start II DNA Polymerase (Thermo Scientific) and primer pairs in back-to-back orientation. Five identical 50 µL PCR reactions were performed simultaneously. The obtained DNA was mixed and subjected to self-ligation by incubation at 22°C for 1 h in the presence of T4 DNA Ligase (Thermo Scientific). Further steps involved inactivation of the ligase by incubation of reaction mix at 65°C for 10 min, destruction of the still present *pET16b::ZmC₄-NADP-ME* by digestion with DpnI (Thermo Scientific) at 37°C for 2 h and DpnI inactivation by incubation at 80°C for 5 min. For the introduction of the F140I substitution the primer pairs "F140I_FW" (5´-[Phos]-TCATGAACACCCTGCG-3´ and "F140I_RV" (5´-[Phos]TCTTCTTAATTTGGAGTTCCTG-3´) were used. The resulting expression plasmid was called *pET16b::F140I_NADP-ME*. For the introduction of the E339A substitution the primer pairs "E339A_FW" (5´-[Phos]-AAAATATAGCAAGAGCCATC-3´) and "E339A_RV" (5´[Phos]-GCAAGCAAATCAAAGGCATTGTG-3´) were used. The resulting expression plasmid was called *pET16b::E339A_NADP-ME*. For the introduction of the Q503E substitution the primer pairs "Q503_FW" (5´-[Phos]-AAGCATATACATGGAGT-3´) and "Q503_RV" (5´[Phos]-CTTCAGCAGTGCATTCA-3´) were used. The resulting expression plasmid was called *pET16b::Q503E_NADP-ME*. For the introduction of the L544F substitution the primer pairs "L544F_FW" (5´-[Phos]-TCGGCCTCGGTCTTGTGATCTC-3´) and "L544F_RV" (5´[Phos]-ATCCAGGGAAAATGTAGGCATTG-3´) were used. The resulting expression plasmid was called *pET16b::L544F_NADP-ME*.

To express ZmnonC₄-NADP-ME, the cDNA sequence coding the mature protein was cloned in pET28 expression vector (Novagen) between the restriction sites *Nde*I and *Xho*I. First, a 1279 bp fragment was amplified from the plasmid *pET32::ZmnonC₄ NADP-ME* [14] using the following primer pairs: "NdeImadI" (5´-CATATGGCCGCGGAGATGGAGCAG-3´) and P4: (5´-GAAAATGTAGGCGTTGTTCGACTGT-3´). The fragment obtained by the digestion of the PCR product with NdeI and BamHI was ligated with a second fragment obtained by the digestion of the plasmid *pET32::ZmnonC₄ NADP-ME* with *BamH*I and *Xho*I and the vector pET-28 digested with *Nde*I and *Xho*I. The resulting expression construct, *pET28::ZmnonC₄-NADP-ME*, was used to introduce the A399E mutation by site-directed mutagenesis essentially as described above, using the following primer pairs: "FOR_nonC4A339E" (5´-AAACAAGATGGCTCTTGCTATATTTTTCAAGCAAATTAAAAGCATTATGATTC-3´) and "REV_nonC4A339E"

(5´-GAATCATAATGCTTTTAATTTGCTTGAAAAATATAGCAAGAGCCATCTTGTTT-3´). The mutation I140F was obtained as described in [14] The oligonucleotides primers that introduced the mutation in the WT protein was "I140F" (5´- CTCCAAATTAAGAAGTTCATGCACAACCT -3´) in which the mutation position is underlined in the oligonucleotide sequence. The fragments containing the mutation was subcloned from the PCR product to the *pET32::ZmnonC4 NADP-ME*, replacing the WT sequence.

Expression and purification of recombinant C4- and nonC4-NADP-ME variants

All expression constructs were transformed into the *E. coli* strain Rosetta™ 2(DE3) (pLysSRARE2) (Merck) for ZmC4-NADP-MEs and in *E. coli* BL21(DE3) cells in the case of ZmnonC4-NADP-MEs. After plasmid preparation, the success of the site-directed mutagenesis procedure was confirmed by sequencing. For the heterologous protein production, transformed cells were grown in 400 mL LB medium in the presence of the appropriate antibiotic in each case at 37°C and 110 rpm until an OD_{600} of 0.6 - 0.8. To induce the expression of the heterologous protein, 1 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) was added to the culture and the cells were incubated 20 h under the same conditions. The cellular culture was harvested at 4.000 x *g* for 15 min and pellets were stored at -20 °C until use.

For protein extraction, pellets were thawed on ice and resuspended in 20 mM Tris-HCl (pH 8.0) containing 500 mM NaCl, 5 mM imidazole, 2 mM phenylmethanesulfonyl fluoride (PMSF) and a spatula-tip amount of lysozyme, sonicated, and centrifuged at 14,000 x g for 20 min at 4°C. The supernatant was used for protein purification using gravity-flow immobilized metal ion chromatography on nickel-nitrilotriacetic acid agarose (Ni-NTA Agarose, Qiagen). The column was pre-equilibrated with 20 mM Tris-HCl buffer containing 500 mM NaCl and 5 mM imidazole. After the supernatant was loaded, the columns were washed in four steps with 500 mM NaCl in 20 mM Tris-HCl (pH 8.0) containing increasing concentrations of imidazole (5, 50, 60, and 65 mM). Protein elution was performed in four steps of 500 µL of 20 mM Tris-HCl, 500 mM NaCl and 300 mM imidazole. The protein eluted in the first fraction was used for further kinetic measurements.

Removal of the His-tag for crystallization of Zm- and SbC4-NADP-ME

For crystallization studies of ZmC4-NADP-ME all elution fractions were pulled together and subjected to concentration and buffer exchange in a Vivaspin 20 centrifugal concentrator (Sartorius AG) using 20 mM Tris-HCl (pH 8.0), 100 mM NaCl and 2 mM CaCl2. Final concentrations of the protein samples were between 10 and 14 mg mL^{-1}. The N-terminal His-Tag of the recombinant enzyme was cleaved off by incubation with factor Xa protease (New England Biolabs), the recognition sequence for which was encoded by the pET16b vector and is located between the His-tag and the ZmC4-NADP-ME sequences. One µg protease was added to 50 µg of ZmC4-NADP-ME and incubated at 23°C for 3h. Cleaved His-Tag and factor Xa protease were eliminated by size exclusion chromatography employing HiPrep 16/60 Sephacryl S-300 High Resolution column (GE Healthcare Life Sciences) pre-equilibrated with 10 mM Tris-HCl (pH 8.0) containing 0.1 M NaCl. Gel filtration was carried out using the ÄKTA prime plus system (GE Healthcare Life Sciences). Elution was carried out using 120 mL of the same buffer at 0.2 mL min^{-1}. The protein eluted as a single peak at the volume expected for the homotetramer. In the case of SbC4-NADP-ME, the His-tag fusion was removed by incubation with 1:20 (thrombin: SbC4-NADP-ME) for 2 hs at 16°C. The cleaved His-tag was eliminated using a similar protocol as described before for ZmC4-NADP-ME, by size exclusion chromatography. In each case, protein fractions were combined and subjected to concentration and buffer exchange using 10 mM Tris-HCl (pH 8.0) containing 5 mM MgCl2 employing a Vivaspin® 20 centrifugal concentrator (Sartorius AG). For further crystallization studies protein concentration of the sample was adjusted to 10 mg mL^{-1} in the case of ZmC4-NADP-ME and 6 mg mL^{-1} for SbC4-NADP-ME.

Zm- and SbC4-NADP-ME crystallization and X-ray data collection

The initial crystallization condition for ZmC4-NADP-ME without His-tag was identified from sparse matrix screening using commercial screens (Molecular Dimension, Suffolk, UK and Nextal, Qiagen, Hilden, Germany). Crystals were grown at 12°C by the sitting drop vapor diffusion method. 0.1µL of the enzyme, at a concentration of 10 mg mL^{-1}, was mixed with 0.1µL of the reservoir solution and equilibrated against 40µL reservoir solution. The most promising condition contained 0.2 M potassium chloride, 0.1 M sodium citrate (pH 5.5) and 37% pentaerythritol propoxylate (5/4 PO/OH) from MemGold I, B11 (Molecular Dimensions, Suffolk, UK) and was optimized with protein at 5 mg mL^{-1} by a grid screen (best diffracting crystals grew in 0.1M potassium chloride, 0.1 M sodium citrate (pH 5.5) and 34% pentaerythritol propoxylate). Crystals appeared within a few days and grew to a size about 50 × 50 × 250 µm^3 in 5 to 7 days. The crystals were fished after overlaying the drop with 1µL

mineral oil and flash frozen in liquid nitrogen for X-ray data collection at 100 K. X-ray diffraction data were collected to 2.2 Å resolution at the ESRF beamline ID23-1 (Grenoble, France).

In the case of SbC4-NADP-ME, initial identification of crystallization conditions was carried out using the sitting-drop vapor diffusion method, with a robotic workstation (Honeybee963®, Isogen Life Science). Sitting drops were set using 400 nL of a 1:1 mixture of protein and mother liquor, equilibrating against 150 µL mother liquors in the reservoir of Greiner plates, both at 4°C and 20°C. The mother liquors screened were from commercial kits: Crystal screens HR2-110 and HR-112 (Hampton Research) and PEG Suite solutions (QIAGEN). The best crystal hit contained 0.2 M KF and 20% PEG 3350 as mother liquor. The crystal growth habit comparing both crystallization temperatures was similar, so we continued working only at 20°C. Initial SbC4-NADP-ME crystals were manually optimized by varying precipitant and protein concentrations using VDX plates (Hampton Research) with a hanging-drop setup. The best crystallization condition was ultimately achieved using 0.2 M KF and 30% PEG 3350 reservoir solution at 20°C, with protein at 5.0 mg/mL in 20 mM Tris pH 8.0, 50 mM NaCl, 2 mM NADP, 40 mM pyruvate and 10 mM $MgCl_2$. Crystals appeared in about 7 days. Cryo-protection was achieved by slowly adding cryo-protection solution (25% (v/v) glycerol, 30% (w/v) PEG 3350, 0.1 M Tris-HCl pH 8.0, 0.2 M $MgCl_2$, 2 mM NADP and 40 mM pyruvate) to the drop in small volumes (~ 5% drop volume step-wise, until reaching >15% glycerol), then rapidly soaked in 100% cryo-protection solution, and flash-cooled in liquid N_2 for storage until data collection. X-ray diffraction data collection was performed with a Micromax007-HF generator (Rigaku) with a rotating Cu anode, Varimax-HF (Rigaku) optics and a Mar345 image plate detector (Mar Research) reaching 2.0 Å resolution (Supplementary Table 1).

Data processing, structure determination and refinement

All data obtained for the ZmC4-NADP-ME and SbC4-NADP-ME were processed with XDS [28] or Mosflm [29] and scaled with aimless [30]. Phasing was performed by MR using chain A from the human mitochondrial malic enzyme (pdb-code: 1QR6 and 1DO8 respectively) as template. The resulting structure was used as starting point for further model building. Both structures was then refined by iterative cycles of manual refinement using Coot [31] and Refmac5 [32] from CCP4 suite [33] or Buster [34]. The structures were deposited in the protein data bank (wwpdb) under the accession codes 5OU5 (ZmC4-NADP-ME) and 6C7N (SbC4-NADP-ME). Rmsd

values were calculated using Superpose from CCP4 suite, interface areas were determined using the PDBePISA online server (http://www.ebi.ac.uk/pdbe/pisa/). All images of the models were prepared using MacPyMOL (Delano, W. L. The PyMOL molecular graphics system, 2002).

Protein quantification and gel electrophoresis procedures

The protein concentration of purified enzymes and other protein-containing samples used for analysis was determined using the Pierce^{TM} BCA Protein Assay Kit (Thermo Fischer).

SDS-PAGE was performed in 12% (w/v) polyacrylamide gels according to Laemmli [35]. Proteins were visualized by staining with Coomassie Brilliant Blue or electroblotted onto a polyvinylidene difluoride (PVDF) membrane (Carl Roth) for subsequent immunological detection. The recombinantly expressed and purified proteins showed in all cases the expected molecular masses of 66 kDa (63.4 kDa of the mature ZmC$_4$-NADP-ME plus 2.5 kDa encoded by the expression vector, and 66 kDa of mature ZmnonC$_4$-NADP-ME after cleavage the His-Tag fusion). For native PAGE, the affinity-purified enzymes were separated on non-reducing and non-denaturing 6% (w/v) polyacrylamide gel at 100 V. The separated proteins were analysed by *in-gel* NADP-ME activity or immunological detection.

In-gel NADP-ME activity assay and immunological detection

In-gel NADP-ME activity assays were performed by incubating the gels in the dark at room temperature in a reaction buffer containing 50 mM Tris-HCl (pH 8.0), 50 mM L-malate, 10 mM MgCl$_2$, 0.5 mM NADP, 0.05% (w/v) nitro blue tetrazolium (NBT) and 3 mM phenazinemethosulfate (PMS).

For immunoblotting assays, the separated proteins were transfer to a PVDF membrane (Roti-PVDF, Carl Roth). The membranes were incubated for at least 1h with a 1:7500 dilution of Anti-His-Tag antibody coupled to the horseradish peroxidase (Anti-His-HRP) (Miltenyi Biotec), or alternatively, with a 1:50 dilution of a rabbit polyclonal antibody raised against an epitope of the ZmC$_4$-NADP-ME located on the solvent-exposed surface of the protein (peptide VWLVDSKGLIVD) (Eurogentec). After washing several times, the membranes were incubated with a 1:2500 dilution of the Goat Anti-Rabbit IgG Antibody HRP-conjugate

(Merck). Proteins were visualized by chemiluminescence with Pierce® ECL Western Blotting Substrate (Thermo Scientific) on a LAS-4000 Mini Luminescent Image Analyzer (GE Healthcare Life Sciences).

Determination of kinetic parameters

NADP-ME activity was determined using a Synergy HT Biotek® Plate Reader system by measuring the formation of NADPH at 340 nm at 25°C. The standard assay medium contained 0.5 mM NADP⁺, 10 mM MgCl₂, 4 mM L-malate, 50 mM Tris-HCl buffer (pH 8.0) and 50-200 ng enzyme per well in a final volume of 200 μL.

The dependence of the activity with the pH of the medium was determined with the standard assay medium using different buffer systems: 50 mM MES (pH 5.5-6.5), 50 mM Tricine-Mops (pH 7.0-7.5), and 50 mM Tris-HCl (pH 7.5-9.0).

The apparent Michaelis constants (K_m) for malate and NADP were determined at pH 8.0 by varying the concentration of one component (NADP concentrations were between 2 and 150 μM and malate concentrations were between 0.001 and 8 mM for C4-NADP-ME, and between 0.001 and 25 mM for nonC4-NADP-ME) while keeping the concentration of the other component constant and saturating. Malate inhibition analysis was performed at pH 7.0.

An extinction coefficient (ε) of 6.22 mM⁻¹ cm⁻¹ for NADPH at 340 nm was employed in the calculations. All kinetic parameters were calculated at least with three batches of independently purified proteins, with each calculation containing a triplicate determination. Data were adjusted to non-lineal regression with Prism 6 (GraphPad Software).

Statistical analysis

For statistical analysis of the kinetic data, a one-way ANOVA was performed using Prism 6 (GraphPad Software). Post-hoc analysis was performed using Dunett's testing procedures.

Secondary structure analysis by circular dichroism

CD analysis of all recombinant versions of ZmC4-NADP-ME were performed with the proteins in 20 mM NaP$_i$ (pH 8.0), 5 mM MgCl₂. Amicon Ultra 0.5 mL centrifugal filters with nominal molecular weight limit of 50 kDa (Merck) were used for buffer exchange procedures. Prior to

the CD measurement, protein concentration was determined at 280 nm in a Jasco V-650 spectrophotometer and was adjusted to 0.16 mg mL^{-1}. The protein concentration c in mg mL^{-1} was calculated with the equation $c = OD_{280} \cdot Mw/(\varepsilon \cdot d)$, in which OD_{280} is the measured optical density at 280 nm, Mw is the molecular weight of the protein in g mol^{-1}, $\varepsilon = 61,770$ M^{-1} cm^{-1} is the molar extinction coefficient of the protein at 280 nm, and $d = 1$ cm is the cell path length.

CD spectra were obtained at 20 °C with a Jasco J-715 spectropolarimeter using a 0.1-cm path length cell and averaging 5 repetitive scans between 250 and 190 nm. Mean residue ellipticity ([Θ]) was obtained by the equation, $[\Theta] = MRW \cdot \Theta/(10 \cdot c \cdot d)$, in which MRW (the mean amino acid residue weight) was calculated as the relation of the protein's molecular weight (Mw) to the number of peptide bonds ($N - 1$; where N is the number of amino acids in the protein chain), d is the cell path (0.1 cm), Θ is the ellipticity measured in millidegrees (machine units), and c is the protein concentration obtained from the previous equation. The protein secondary structural composition was estimated from the experimental CD data using the CONTIN-LL method [36,37] at the Dichroweb online server [38] (http://dichroweb.cryst.bbk.ac.uk) [39]. Prism Software (http://www.graphpad.com/scientific-software/prism/) was used for visualization of the deconvolution by Dichroweb.

Quaternary structure analysis by analytical ultracentrifugation

Sedimentation velocity experiments were carried out with ZmC4-NADP-ME WT and F140I in 10 mM Tris-HCl at pH 8.0 and 5 mM MgCl$_2$ using a Beckman Optima XL-A analytical ultracentrifuge. Amicon Ultra 0.5 mL centrifugal filters with nominal molecular weight limit of 50 kDa (Merck) were used for buffer exchange procedures. Prior to the centrifugation, protein concentration was adjusted to 0.6 mg mL^{-1}. Samples (230 µg) and buffer solutions (400 µL) were loaded into aluminium double sector centrepieces separately and built up in a Beckman An-50 Ti rotor. Experiments were performed at 20°C and 35,000 rpm. Protein samples were monitored at 280 nm in a continuous mode with a radial resolution of 0.003 cm. In time intervals of about 2 min scans of the radial concentration profile were collected until the protein was fully sedimented. Data were analysed using the ls-g*(s) model in the software package SEDFIT [40]. For data analysis, a resolution of 0.1 S with a confidence level (F-ration) of 0.683 was chosen for the appropriate s-value range within 0 to 100.0 S in case of the WT

and 0 to 500.0 S in case of the F140I variant. Density and viscosity of the solvent had been calculated with the software Sednterp (vs 20130813 beta, supported by Biomolecular Interaction Technologies Center at the University of New Hampshire) from tabulated values: $\rho = 0.99885$ g cm^{-3} and $\eta = 0.01006$ g cm^{-1} s^{-1}. The protein partial specific volume applied for WT and mutants was 0.7387 cm³ g^{-1}. Graphic output was generated by Gussi (Version 1.2.1) [41] and the final sedimentation coefficient distribution was normalized based on the maximum peak height.

AKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

AB and C-CW produced the ZmC$_4$-NADP-ME recombinant proteins and obtained kinetic and structural data. AB produced protein crystals. AH planned crystallization studies of ZmC$_4$-NADP-ME, collected and analysed structural data and solved the structure. CEA, FT, and AB performed and analysed crystallization studies of SbC$_4$-NADP-ME. CEA and MS produced the ZmnonC$_4$-NADP-ME recombinant proteins and obtained kinetic and structural data. C-CW and MJL designed the algorithm to identify strictly differentially conserved amino acid residues and performed the bioinformatic analyses. TZ and LN-S performed and analysed CD and AUZ analysis. VGM conceived and led the project and together with MFD designed and supervised the work and analysed data. AB, AH, CEA, TZ, MFD, MJL, and VGM wrote the manuscript and generated the figures.

REFERENCES

1   Amthor, J. S. From sunlight to phytomass: on the potential efficiency of converting solar radiation to phyto-energy. *New Phytol.* **188,** 939-959 (2010).

2   Osborne, C. P. & Sack, L. Evolution of C4 plants: a new hypothesis for an interaction of CO2 and water relations mediated by plant hydraulics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **367,** 583-600 (2012).

3   Sage, R. F., Sage, T. L. & Kocacinar, F. Photorespiration and the Evolution of C-4 Photosynthesis. *Annu. Rev. Plant Biol.* **63,** 19-47 (2012).

4   Sage, R. F., Christin, P. A. & Edwards, E. J. The C(4) plant lineages of planet Earth. *J. Exp. Bot.* **62,** 3155-3169 (2011).

5   Emms, D. M., Covshoff, S., Hibberd, J. M. & Kelly, S. Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C4 Species. *Mol. Biol. Evol.* **33,** 1796-1806 (2016).

6   Maier, A., Zell, M. B. & Maurino, V. G. Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C(4) and C(3) photosynthesis. *J. Exp. Bot.* **62,** 3061-3069 (2011).

7   Maurino, V. G., Saigo, M., Andreo, C. S. & Drincovich, M. F. Non-photosynthetic 'malic enzyme' from maize: a constituvely expressed enzyme that responds to plant defence inducers. *Plant Mol. Biol.* **45,** 409-420 (2001).

8   Saigo, M. *et al.* Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Mol. Biol.* **55,** 97-107 (2004).

9   Tausta, S. L., Coyle, H. M., Rothermel, B., Stiefel, V. & Nelson, T. Maize C4 and non-C4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Mol. Biol.* **50,** 635-652 (2002).

10   Christin, P. A., Samaritani, E., Petitpierre, B., Salamin, N. & Besnard, G. Evolutionary insights on C4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome biology and evolution* **1,** 221-230 (2009).

11   Saigo, M., Alvarez, C. E., Andreo, C. S. & Drincovich, M. F. Plastidial NADP-malic enzymes from grasses: unraveling the way to the C4 specific isoforms. *Plant Physiol. Biochem.* **63,** 39-48 (2013).

12   Ashton, A. R. NADP-malic enzyme from the C4 plant Flaveria bidentis: nucleotide substrate specificity. *Archives of biochemistry and biophysics* **345,** 251-258 (1997).

13   Iglesias, A. A. & Andreo, C. S. Kinetic and Structural Properties of NADP-Malic Enzyme from Sugarcane Leaves. *Plant Physiol.* **92,** 66-72 (1990).

14   Detarsio, E., Wheeler, M. C., Campos Bermudez, V. A., Andreo, C. S. & Drincovich, M. F. Maize C4 NADP-malic enzyme. Expression in Escherichia coli and characterization of site-directed mutants at the putative nucleoside-binding sites. *J. Biol. Chem.* **278,** 13757-13764 (2003).

15   Detarsio, E., Alvarez, C. E., Saigo, M., Andreo, C. S. & Drincovich, M. F. Identification of domains involved in tetramerization and malate inhibition of maize C4-NADP-malic enzyme. *J. Biol. Chem.* **282,** 6053-6060 (2007).

16   Werdan, K., Heldt, H. W. & Milovancev, M. The role of pH in the regulation of carbon fixation in the chloroplast stroma. Studies on CO2 fixation in the light and dark. *Biochim. Biophys. Acta* **396,** 276-292 (1975).

17   Fahnenstich, H. *et al.* Alteration of organic acid metabolism in Arabidopsis overexpressing the maize C(4)NADP-malic enzyme causes accelerated senescence during extended darkness. *Plant Physiol.* **145,** 640-652 (2007).

18 Zell, M. B. *et al.* Analysis of Arabidopsis with highly reduced levels of malate and fumarate sheds light on the role of these organic acids as storage carbon molecules. *Plant Physiol.* **152,** 1251-1262 (2010).

19 Roalson, E. H. in C4 photosynthesis and related CO2 concentrating mechanisms. (ed A.S. Raghavendra, Sage, R.F.) 319–338 (Kluwer, 2011).

20 Saigo, M. *et al.* Biochemical approaches to C4 photosynthesis evolution studies: the case of malic enzymes decarboxylases. *Photosynth Res* **117,** 177-187 (2013).

21 Hsieh, J. Y., Chen, S. H. & Hung, H. C. Functional roles of the tetramer organization of malic enzyme. *J. Biol. Chem.* **284,** 18096-18105 (2009).

22 Izui, K., Matsumura, H., Furumoto, T. & Kai, Y. Phosphoenolpyruvate carboxylase: a new era of structural biology. *Annu. Rev. Plant Biol.* **55,** 69-84 (2004).

23 Maurino, V. G. & Weber, A. P. Engineering photosynthesis in plants and synthetic microorganisms. *J. Exp. Bot.* **64,** 743-751 (2013).

24 Covshoff, S. & Hibberd, J. M. Integrating C4 photosynthesis into C3 crops to increase yield potential. *Curr. Opin. Biotechnol.* **23,** 209-214 (2012).

25 Hibberd, J. M., Sheehy, J. E. & Langdale, J. A. Using C-4 photosynthesis to increase the yield of rice - rationale and feasibility. *Curr. Opin. Plant Biol.* **11,** 228-231 (2008).

26 Kajala, K. *et al.* Strategies for engineering a two-celled C(4) photosynthetic pathway into rice. *J. Exp. Bot.* **62,** 3001-3010 (2011).

27 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343-345 (2009).

28 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66,** 125-132 (2010).

29 Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67,** 271-281 (2011).

30 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69,** 1204-1214 (2013).

31 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486-501 (2010).

32 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67,** 355-367 (2011).

33 Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50,** 760-763 (1994).

34 Bricogne, G. *et al.* BUSTER version 2.9. . *Cambridge, United Kingdom: Global Phasing Ltd.* (2009).

35 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680-685 (1970).

36 van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R. & Groen, F. C. Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem.* **191,** 110-118 (1990).

37 Provencher, S. W. & Glockner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20,** 33-37 (1981).

38 Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287,** 252-260 (2000).

39 Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32,** W668-673 (2004).

40 Schuck, P. & Rossmanith, P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **54,** 328-341 (2000).

41   Brautigam, C. A. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol.* **562,** 109-133 (2015).
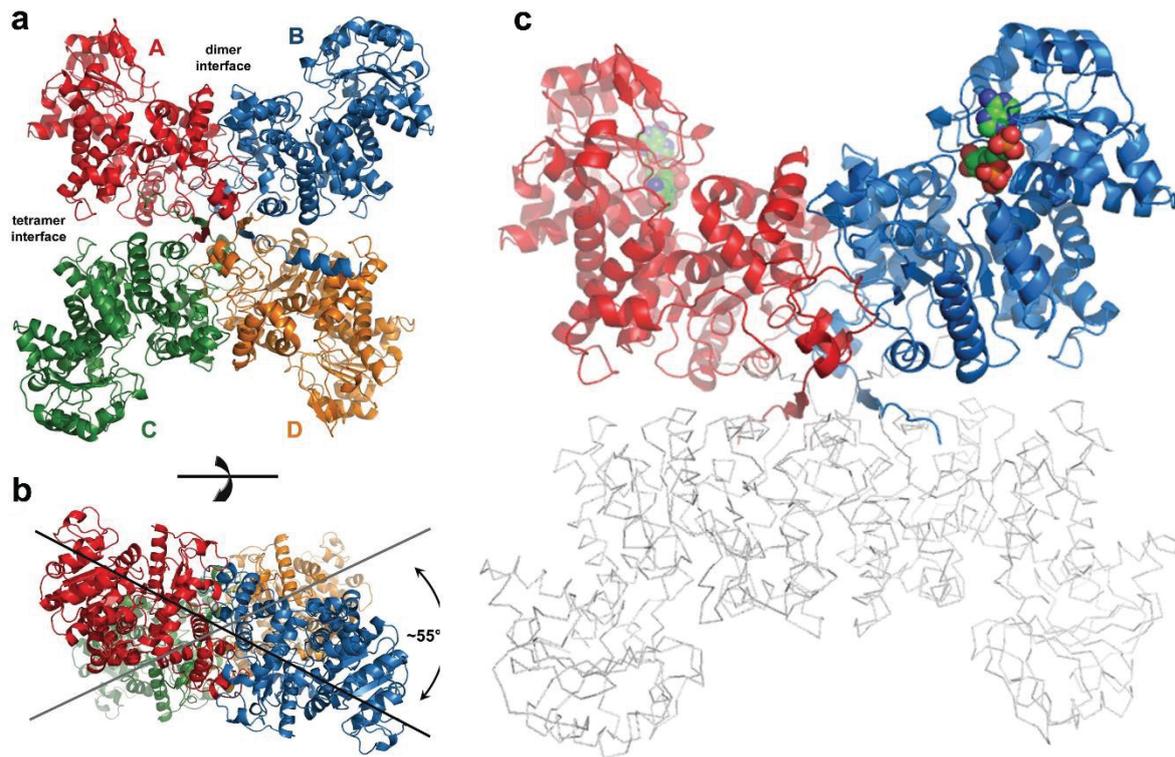
Figure 1. Quaternary organization of C₄-NADP-ME. a, Cartoon illustration of the tetrameric assembly of ZmC₄-NADP-ME. Monomer A (red) and monomer B (blue) represent the dimer, likewise monomer C (green) and monomer D (orange). **b**, Top view on the ZmC₄-NADP-ME, displaying the tilted arrangement of the dimers relative to each other. **c**, Cartoon and ribbon representation of SbC₄-NADP-ME. Monomer A (red) and B (blue) with bound cofactor NADP (colored spheres). Large parts of chain C and D are not well resolved by electron density; therefore, we depicted these chains as gray ribbons only for clarity.
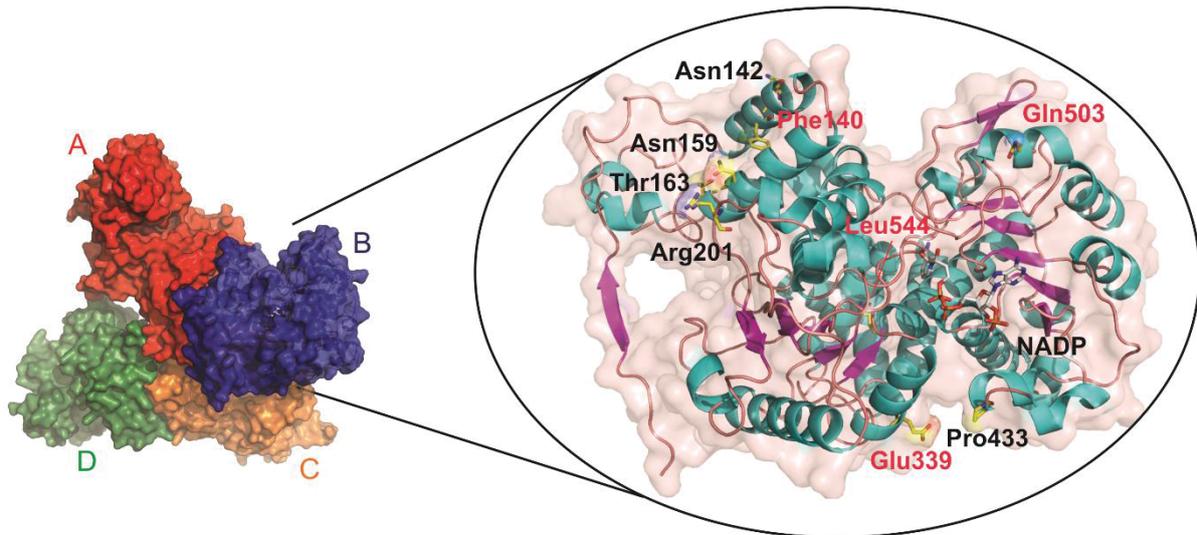
Figure 2. Candidate amino acids for functional analysis identified using Zm- and SbC4-NADP-ME crystal structures. The C4-NADP-ME tetramer is shown on the left (color coding as in Fig. 1a). A detailed view of monomer B is given on the right, showing all candidate amino acids for the mutational analyses. The four amino acids also strictly differentially conserved in *S. italica* and examined in substitution mutants in this study are labeled in red. F140, N142, N159, and T163 represent positions comparable to the fumarate binding site (allosteric activator) in human malic enzyme (pdb code 1PJ3); R201 is in close proximity to N159 and points towards the dimer interface; E339 is located in helix αB4 which in turn interacts with the cofactor via N331 as observed in SbC4-NADP-ME (Supplementary Fig. 3); the rigid P433 is close to K435, which binds the 2´-phosphate in NADP; Q503 is part of helix αC5 and may stabilize the loop containing L490 and N492 which both bind NADP in SbC4-NADP-ME (Supplementary Fig. 3); L544 is part of helix αB5, which interacts with its N-terminal end with the cofactor NADP via S535 an N537 in SbC4-NADP-ME (Supplementary Fig. 3). The position of the cofactor NADP indicates the active site.

a

ZmC4-NADP-ME

Coomassie   Immunoblot

b

c   **d**   **e**

f

ZmC4-NADP-ME	WT	F140I	E339A	Q503E	L544F
K_m malate (mM)	0.23 ± 0.02	0.47 ± 0.13[*]	0.47 ± 0.17[*]	0.46 ± 0.12[*]	0.58 ± 0.04[**]
K_m NADP (µM)	14.9 ± 2.2	8.2 ± 2.8[*]	15.9 ± 2.2	12.8 ± 3.4	15.9 ± 2.8
k_{cat} (s^{-1})	28.1 ± 5.1	3.6 ± 1.3[****]	28.7 ± 3.3	29.8 ± 5.5	27.4 ± 1.8
k_{cat}/K_m (s^{-1} mM^{-1})	122.2	7.7	61.1	64.8	47.2
K_i malate [#](mM)	9.1 ± 2.6	15.8 ± 6.6	–	16.7 ± 2.4	13.4 ± 8.0
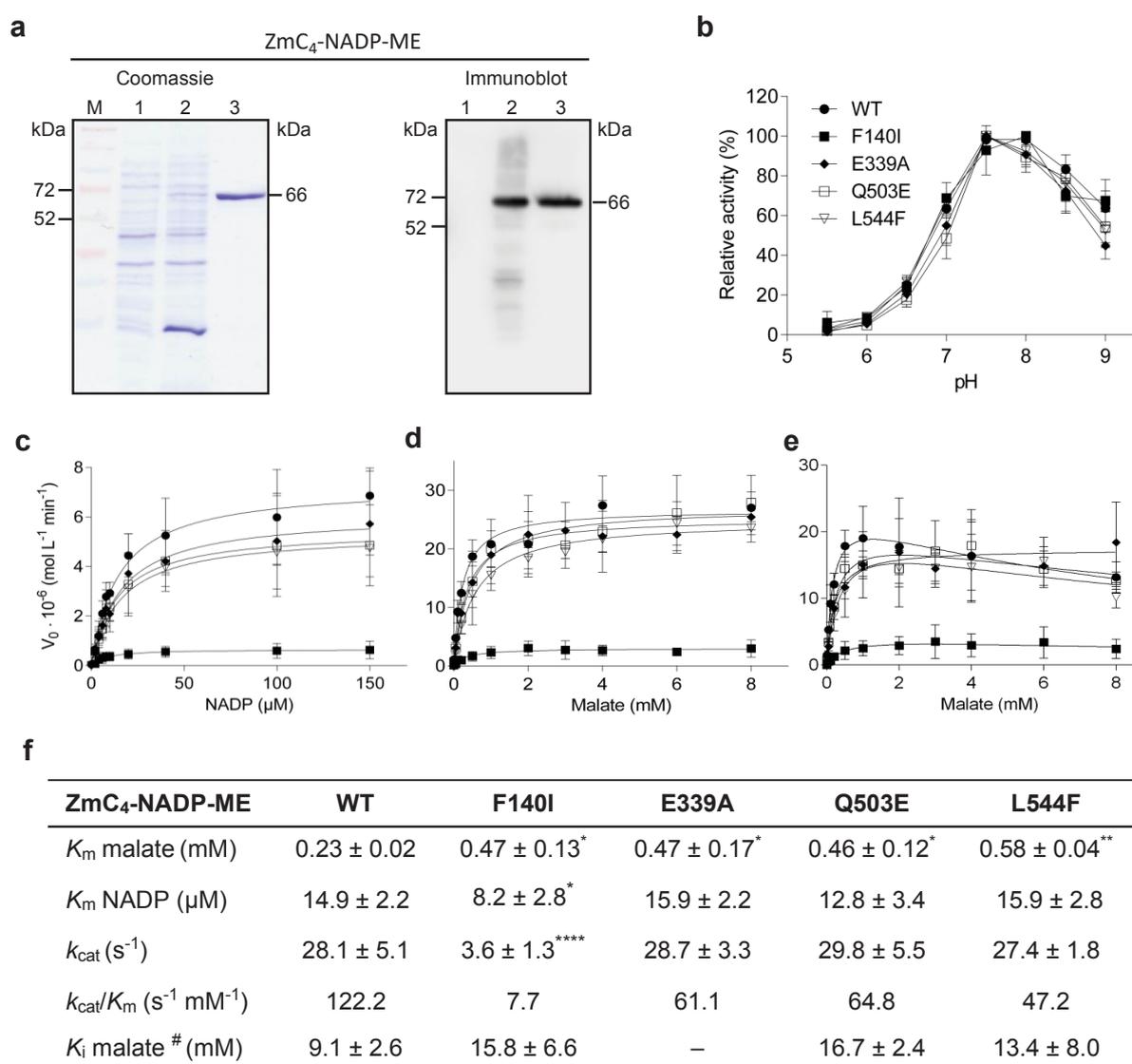
Figure 3. Comparison of ZmC4-NADP-ME variants shows that the identified amino acids contribute to the kinetic differences between C4- and nonC4 isoforms. a, SDS-PAGE stained with Coomassie (on the left) and analysed by immunoblot (on the right) – using anti-His-HRP conjugate antibodies – of different protein fractions during the isolation of recombinant wild-type (WT) ZmC4-NADP-ME. M, molecular weight markers - Spectra^{TM} Multicolor Broad Range Protein Ladder (ThermoFisher Scientific); 1, 10 µg of non-induced cell culture lysate; 2, 10 µg of cell culture lysate 16 h after induction of protein production; 3, affinity-purified ZmC4-NADP-ME; 1.5 µg for protein staining and 0.5 µg for immunoblot. The size of the recombinant ZmC4-NADP-ME (66 kDa) is indicated on the left. **b**, Dependence of the activity of ZmC4-NADP-ME WT and the four engineered enzyme versions with the pH of the reaction media. The values represent the mean ± standard deviation of at least three

independent enzyme preparations, each measured in triplicate. For each enzyme version the highest activity measured (V_0 in μmol mg^{-1} min^{-1} WT: 25.25, F140I: 1.42, E339A: 24.05, Q503E: 22.82, and L544F: 21.30) was set to 100%. **c**, ZmC$_4$-NADP-ME activity at varying amounts of NADP at pH 8.0 using 50 ng enzyme per assay. **d**, ZmC$_4$-NADP-ME activity at varying amounts of malate at pH 8.0 using 200 ng enzyme per assay. **e**, ZmC$_4$-NADP-ME activity at varying amounts of malate at pH 7.0 using 200 ng enzyme per assay. **f**, Kinetic parameters of recombinant ZmC$_4$-NADP-ME WT and the four mutated versions. Kinetic data was best fitted by nonlinear regression analysis. The values represent the mean $\pm$ standard deviation of at least three independent enzyme preparations, each measured in triplicate. The parameters were determined at pH 8.0 with the exception of the *Ki* malate, which was determined at pH 7.0 ($^{\#}$). –, no inhibition by malate was measured. Asterisks (*) indicate that the value is statistically significantly different from the corresponding WT value at the 0.05 level (*), 0.01 level (**), or 0.0001 (****) level.

a

ZmnonC$_4$-NADP-ME	WTnon-C$_4$	I140F	A339E
K$_m$ malate (mM)	0.43 ± 0.06	0.85 ± 0.13**	0.27 ± 0.03*
K$_m$ NADP (mM)	23.0 ± 3.0	38.0 ± 4.6**	15.0 ± 3.0*
k$_{cat}$ (s^{-1})	26.3 ± 0.8	39.6 ± 1.0***	32.1 ± 0.8***
K$_{cat}$/K$_m$ (s^{-1} mM^{-1})	62.6	46.6	118.9
K$_i$ malate [#] (mM)	-	-	19.4 ± 2.3
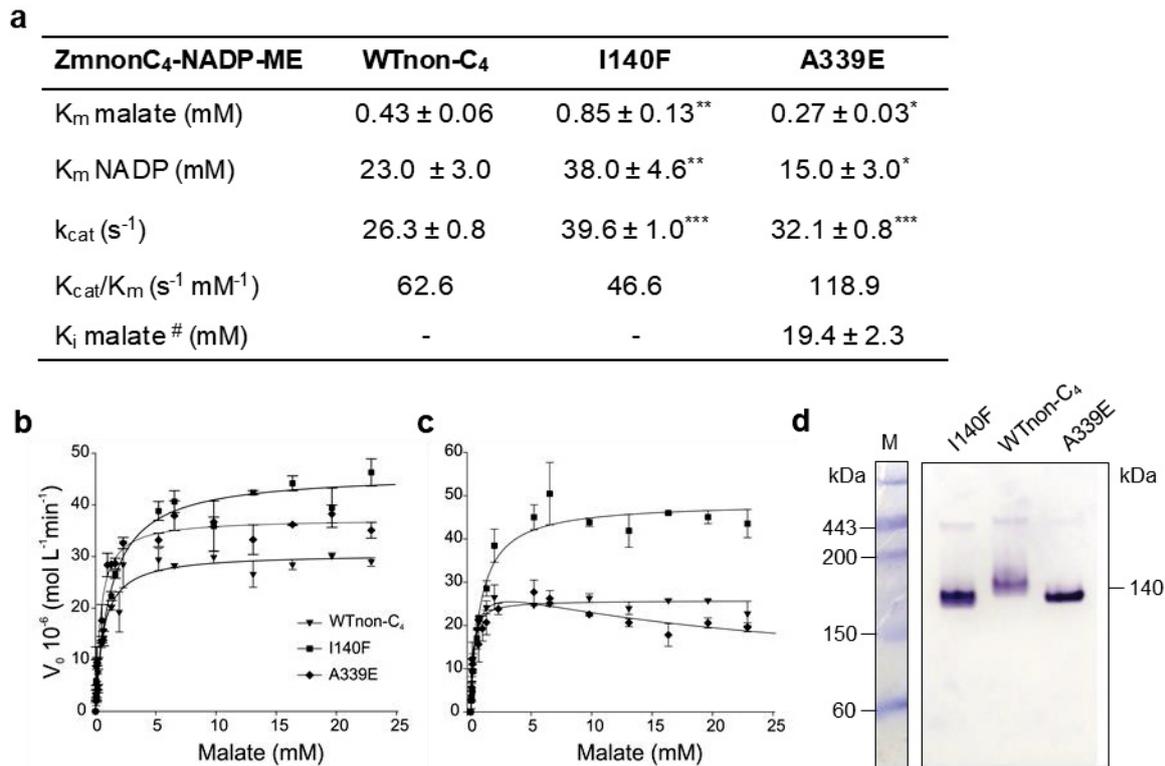
Figure 4. Comparison of ZmnonC$_4$-NADP-ME variants confirms pivotal roles of the evolutionary I140F and A339E substitutions. a, Kinetic parameters of recombinant ZmnonC$_4$-NADP-ME WT and the mutant variants A339E and I140F. Kinetic data were best fitted by nonlinear regression analysis. The values represent the mean ± standard deviation of at least three independent enzyme preparations, each measured in triplicate. The parameters were determined at pH 8.0 with the exception of the Ki for malate, which was determined at pH 7.0 ([#]). –, no inhibition by malate was determined. Asterisks (*) indicate that the value is statistically significantly different form the corresponding WT value at the 0.05 level (*), 0.01 level (**), or 0.001 (***) level. **b**, Representations of ZmnonC$_4$-NADP-ME activity at varying amounts of malate at pH 8.0 using 200 ng enzyme per assay. **c**, Representations of ZmnonC$_4$-NADP-ME activity at varying amounts of malate at pH 7.0 using 200 ng enzyme per assay. **d**, Native PAGE of recombinant ZmnonC$_4$-NADP-ME (WT nonC$_4$) and the I140F and A339E mutant versions (1.0 µg protein per lane) followed by *in gel* NADP-ME activity assay. Coomassie-stained native molecular weight markers – MWGF1000-1KT (Sigma-Aldrich) – are shown to the left.
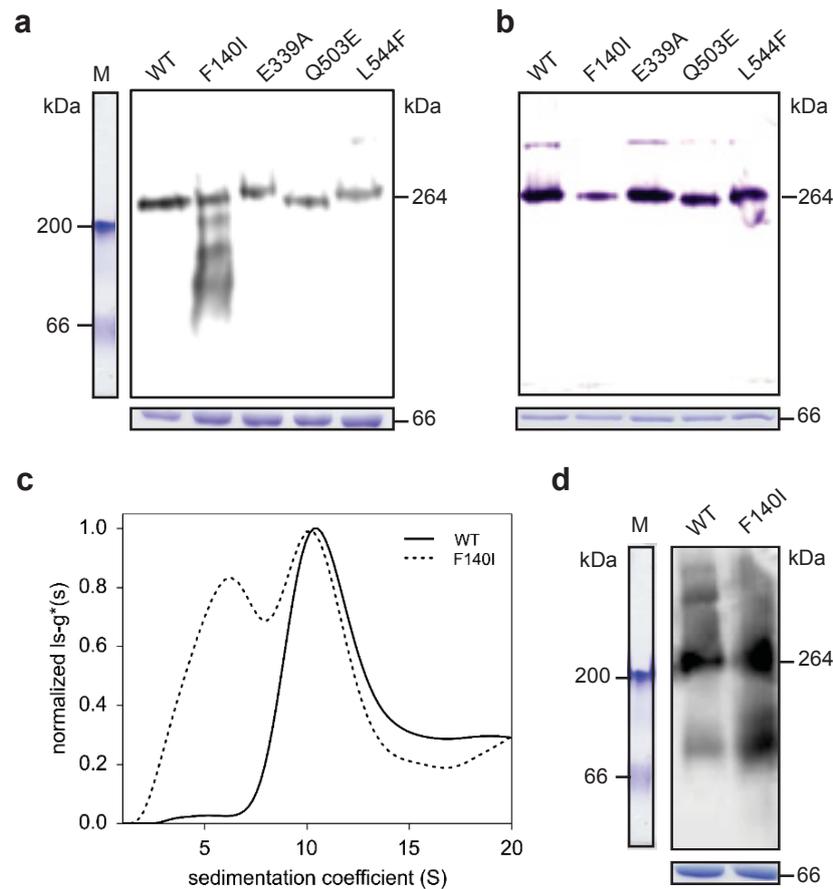
Figure 5. Phe140, but none of the other strictly differentially conserved residues contributes to tetramer formation. a, Native PAGE of recombinant ZmC4-NADP-ME WT and the four mutated versions (3.0 µg protein per lane) followed by immunoblot analysis using anti-ZmC4-NADP-ME antibodies. Coomassie-stained native molecular weight markers - Bovine serum albumin (66 kDa) and ß-Amylase (200 kDa) (Sigma-Aldrich) - are shown to the left. **b**, Native PAGE of recombinant ZmC4-NADP-ME WT and the mutated versions (1.0 µg protein per lane) followed by *in gel* NADP-ME activity assay. **c**, Zoom-in for the sedimentation coefficient distribution between 0 and 20 S of recombinant ZmC4-NADP-ME WT and the F140I mutant at pH 8.0. Data was fitted with the ls-g*(s) model in the software package SEDFIT and normalized based on the maximum peak height. **d**, Native PAGE of recombinant ZmC4-NADP-ME WT and the F140I mutant (10 µg protein per lane) analysed by immunoblot using anti-ZmC4-NADP-ME antibodies. The proteins were incubated at pH 7.0 over 16 h previous to the electrophoresis. The expected molecular weight of the ZmC4-NADP-ME homotetramer (264 kDa) is indicated on the right. Coomassie-stained loading controls (ZmC4-NADP-ME, 66 kDa) are shown below the native gels. Coomassie-stained native molecular weight markers are shown to the left.
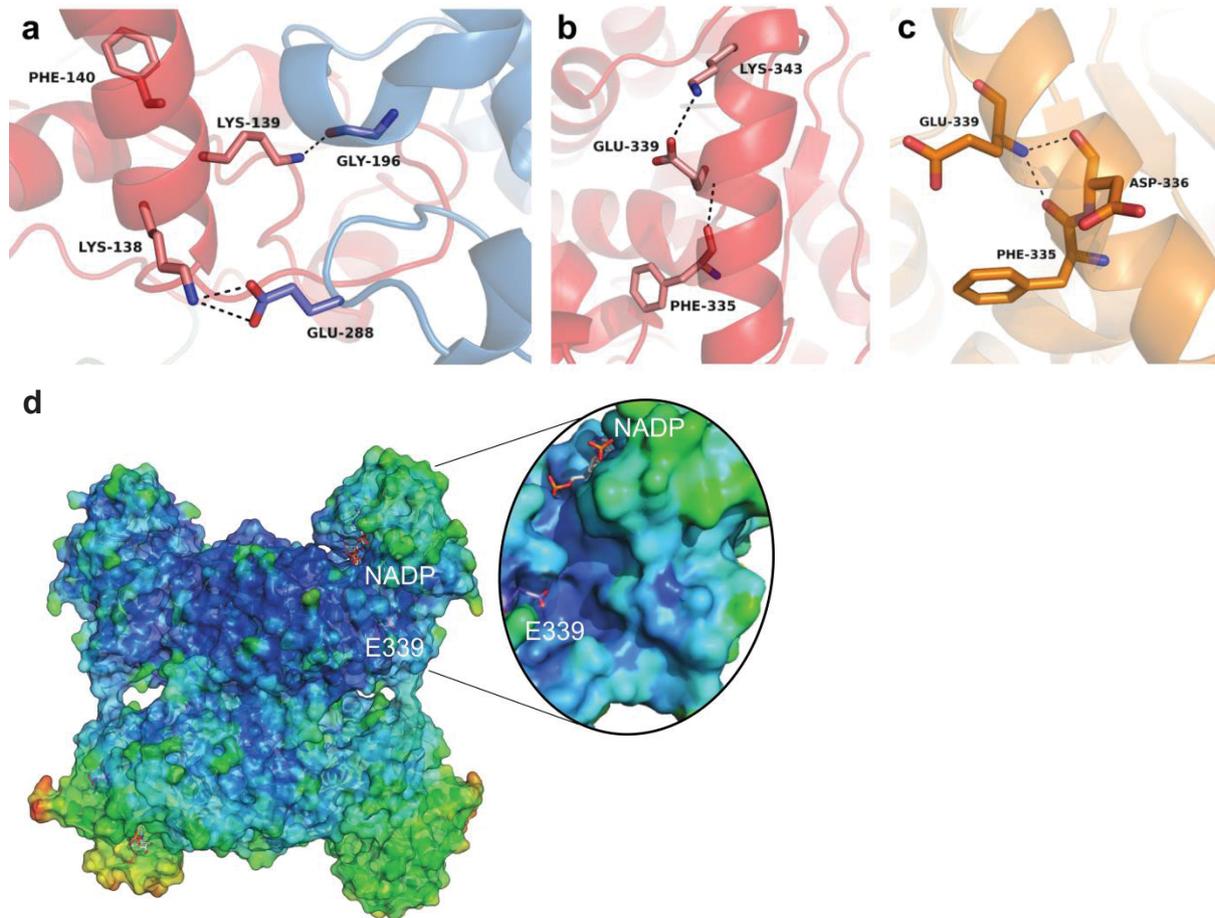
Figure 6. Positions and interactions of Phe140 and Glu339 in the C$_4$-NADP-ME structure. **a**, In monomer A Lys139 and Lys138 next to Phe140 are involved in dimer interactions with Gly196 and Glu288 from the adjacent monomer B in ZmC$_4$-NADP-ME. **b**, Glu339 in all four chains interacts with Phe335, in chain A additionally with Lys343 in ZmC$_4$-NADP-ME. **c**, Glu339 in chain D interacts with Asp336 in ZmC$_4$-NADP-ME. Colour code as in Fig. 1, red: monomer A, blue: monomer B, yellow: monomer D. **d**, Surface representation of SbC$_4$-NADP-ME showing the position of E339, which allows delimiting the malate allosteric binding site. The surface is shown as the b-factor colors of each residue.

Supplementary information

Crystal structure of Zm- and SbC$_4$-NADP-ME: overall structure and assembly

Zm- and SbC$_4$-NADP-ME assemble as homo-tetramer (dimer of dimers). The overall fold is here described for monomer A according to the domain definition of the human mitochondrial NAD(P)-ME (1QR6, Protein Data Bank) (Xu et al., 1999) (Supplementary Fig. 1): Domain A (84-202) consists of one β-strand (βA1) followed by six helices (αA1-αA6). Interestingly, only in monomer B the first N-terminal stretch (71-84) is clearly visible in the electron density, forming an additional helix (αA0). Domain B is composed of two segments of the protein chain (residues 203-349 and 535-607) with a four-stranded parallel β-sheet (βB1, βB3, βB4, βB5) flanked by six helices on both sides (αB1, αB2, αB3´, αB3, αB4, αB5), and an antiparallel helix bundle built-up of αB6, αB7 and αB8. Contrary to the human NAD(P)-ME, βB2 is missing in the ZmC$_4$-NADP-ME structure (residues 229-233); this part of the sequence is instead arranged as an elongated loop between αB1 and αB2. Domain C (350-534) contains a central parallel β-sheet (βC2, βC1, βC4, βC5, βC6) surrounded by eight helices (αC0, αC1, αC2´, αC2´´, αC2, αC3, αC4, αC5). Two short β-strands (βC6´and βC7´) form a hairpin structure at the C-terminus of domain C. The dinucleotide-binding signature motif (Rossmann fold) is located between βC1 and αC1 (383-GAGEAG-388). Domain D (611-636) is comprised of one α-helix followed by a long tail ending up close to αA2 from subdomain A.

The overall structure of SbC$_4$-NADP-ME is highly similar to that of ZmC$_4$-NADP-ME (superposition of SbC$_4$-NADP-ME chain A and ZmC$_4$-NADP-ME chain C: RMSD 0.696 Å over 523 Cα atoms; SbC$_4$-NADP-ME chain B and ZmC$_4$-NADP-ME chain D: RMSD 0.422 Å over 515 Cα atoms). Only parts of chain C and D, which are close to the monomers A and B, are well defined; the sections from aa 360-530 in chains C and D show either no or very poor electron density, reflecting flexible parts in the crystal). We therefore only refer to chains A and B of SbC$_4$-NADP-ME.

The ZmC$_4$-NADP-ME tetramer is composed of two dimers (monomers A and B, and monomers C and D, respectively). The dimers are tilted relative to each other at an angle of ~55° (Fig. 1). The dimer interface areas between monomers A and B and between monomers C and D each amount to ~2300 Å^2. The tetrameric interface between the two dimers spans a very similar area of ~2400 Å^2, divided equally between the two subunit interfaces A-C and B-D. SbC$_4$-NADP-ME is likewise arranged as a tetramer, but due to the insufficiently defined electron density for chains C and D we determined only the interface area between monomer A and B, which amounts to ~2400 Å^2.
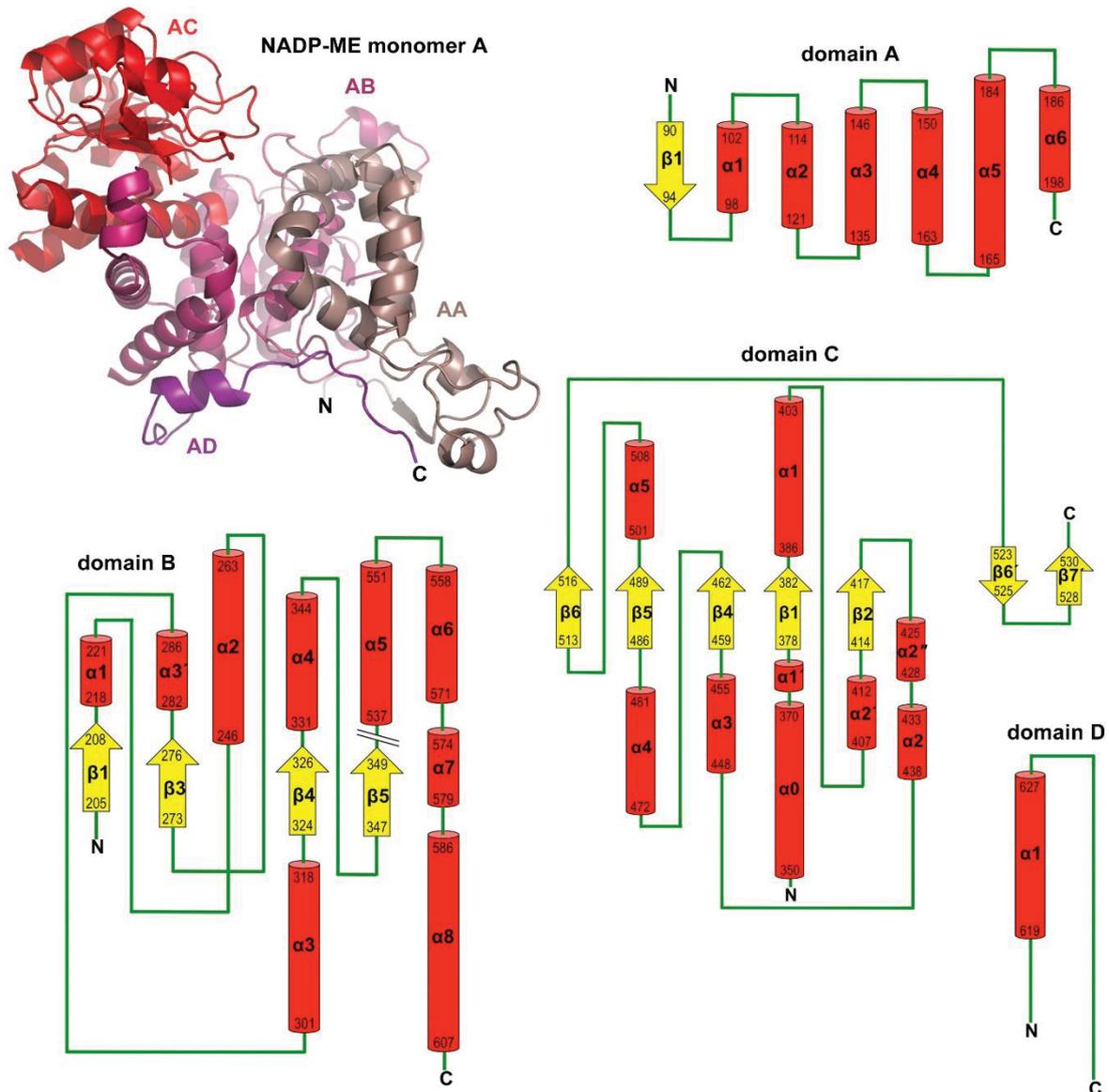
Supplementary Fig. 1. Representation of the domains of monomer A in Zm- and SbC₄-NADP-ME. Top left: Cartoon with domain A in dark salmon, B in pink, C in red and D in purple. Top right and bottom: Topographical schemes of domains A, B, C and D of monomer A. α-helices are shown as red tubes, β-strands as yellow arrows, loops as green lines. The first and last residue positions are indicated at the ends of each secondary element.
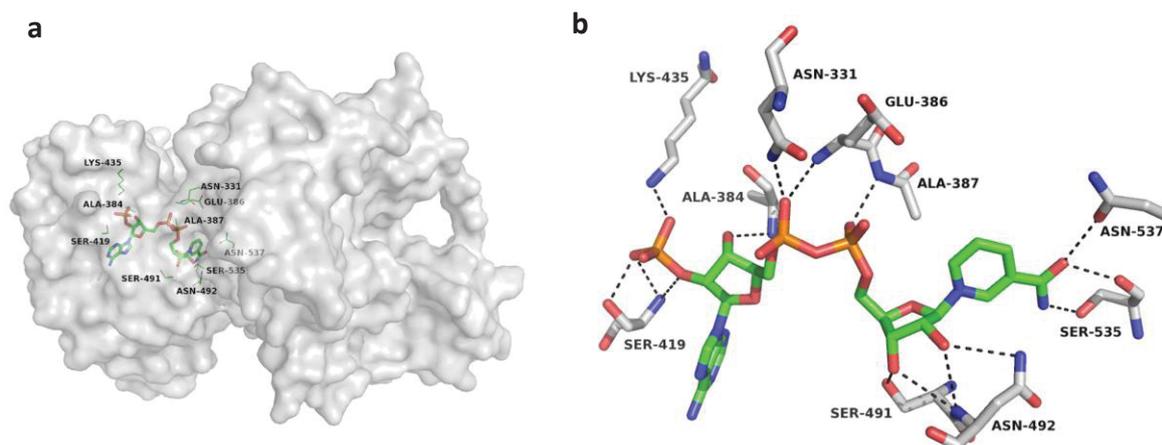
Supplementary Fig. 2. Amino acids involved in cofactor binding of C$_4$-NADP-ME. a, Monomer A from SbC$_4$-NADP-ME with the bound cofactor NADP (represented by sticks). Residues involved in cofactor binding are depicted as lines with labels. **b,** Hydrogen bonding pattern in the SbC$_4$-NADP-ME cofactor binding site (distances are omitted for clarity and are listed in Supplementary Table 3).
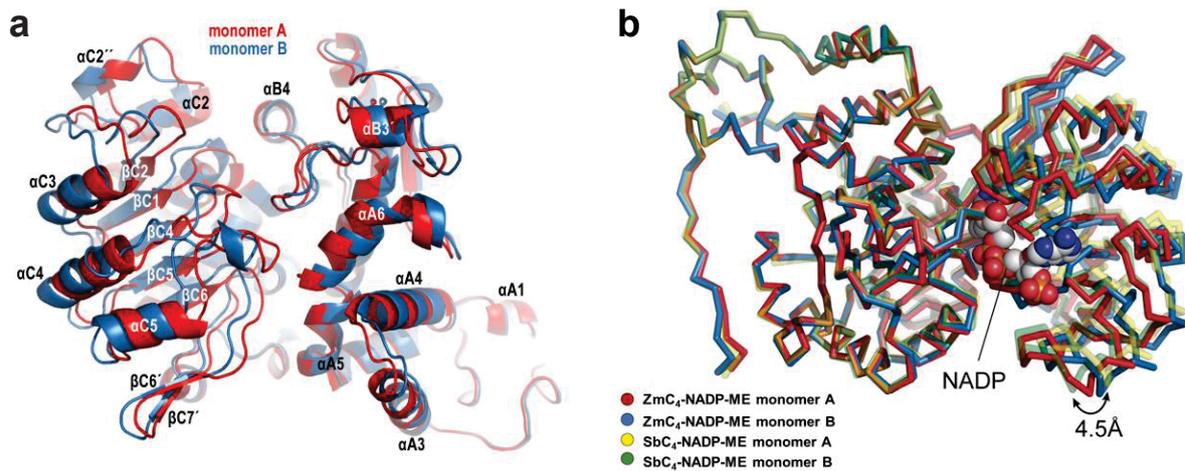
Supplementary Fig 3. Superposition of C₄-NADP-ME monomers indicate different conformations. a, Superposition of monomer A (red) and B (blue) from ZmC₄-NADP-ME, in which the main secondary structure elements from both monomers are labelled. While most of the backbone from domains A and B maintain the same spatial position, domain C in monomer B is shifted outwards, thereby adopting an "open" conformation. **b**, Superposition of the monomers A and B from ZmC₄-NADP-ME and from SbC₄-NADP-ME with bound cofactor NADP. The "open" (monomer B) and "closed" (monomer A) conformation in SbC₄-NADP-ME is comparable to that of ZmC₄-NADP-ME, showing a global movement of parts of the structure by 4.5 Å.

```
                          20              40              60              80
                           |               |               |    ↓          |
ZmC4-NADP-ME   MLSTRTAAVAASASPASPWKLGG-RSE--GGASCDGCRTYRNTLRRRAAPAKVRALPPRRVDAVAMVSNAETETEKEQEE 77
SbC4-NADP-ME   .I.A.AT...-........R..G...--.AG.....S...K.V.....A.......K..E...IG.A....EVVV. 77
SiC4-NADP-ME   ...A..-..SA...AV.GSK----.G.........ESV.....TVR.H.AEQ...Q...VGTA...Q------. 69
ZmnonC4-NADP-ME .I.A....--..--..L..Q..-.......G.............-..A........S.A.E..K.G.A..M.Q.--EA. 70
SbnonC4-NADP-ME .I.A.....--SP.....A..R..G...VG..S............S..A.....V.....E....G.A....K...EQQ 78
Os-NADP-ME     ...A.A..T..A.AASPL..R.EGG.S-GS.SG.TS..E----V.....AVR...AA...E....E.A....KK----. 71

                          100             120             140            160
                           |               |               |          *  |*
ZmC4-NADP-ME   AAAA---------------SEELPVMPWATSVASGYTLLRDPHHNKGLAFTEEERDGHYLRGLLPPAVLSQELQIKKFMN 142
SbC4-NADP-ME   V..---------------N.............................R.....K..A.........V........ 142
SiC4-NADP-ME   GV..ADGGVVDPYAED----.........Y...................K..A.........V..V....H 145
ZmnonC4-NADP-ME V...AAGVVVEDHYGEDSAT....IL...F.................M.K..A.H...........VP......I.H 150
SbnonC4-NADP-ME EVV.GAGGLVEDHYDEDSAA..V.I....F.................K..A.H...........VP.Q..V..I.H 158
Os-NADP-ME     EV..AGGGV------EDMAT..V..T...F.................S.K..A...........V..D..V..I.H 145

                  180          *   *          200              220              240
                   |   *    ***         |               |   **          |
ZmC4-NADP-ME   TLRQYQTPLQRYIAMMNLQETDERLFYKLLIDNVVELLPFVYTPTVGEACQKYGSIFGRPQGLYVSLKDKGKVLEVLRNW 222
SbC4-NADP-ME   N.....L.I.C.M..........E......Y...........................R......... 222
SiC4-NADP-ME   N.....L....M...D...RN.....R.......E................RQ.........R...R........ 225
ZmnonC4-NADP-ME N.....V....M...D...RN........E...V................RQ.........R......... 230
SbnonC4-NADP-ME N.....V....M...D...RN........E...V................RQ.........R......... 238
Os-NADP-ME     N....SV.....M...D...RN........E...V................RQ.........D...... 225

                          260             280             300             320
                           |               |    *          |               |
ZmC4-NADP-ME   PHRNIQVICVTDGERILGLGDLGCQGMGIPVGKLALYTALGGVDPSVCLPITIDVGTNNEKLLNDEFYIGLRQKRATGEE 302
SbC4-NADP-ME   ....V.................................A.....................R... 302
SiC4-NADP-ME   .Q.D.........G.......A...............R..A.............E..............K. 305
ZmnonC4-NADP-ME .....................S...............R..A............................. 310
SbnonC4-NADP-ME ....E................S...............R..A............................. 318
Os-NADP-ME     .E......V...........S...............R..A............Q..........R....K. 305

                  340          *           360             380          *      400
               *            *              |               |              *    |
ZmC4-NADP-ME   YDELIEEFMSAVKQFYGEKVLIQFEDFANHNAFDLLEKYSKSHLVFNDDIQGTASVVLAGLLAALKMVGGTLAEQTYLFL 382
SbC4-NADP-ME   ......M....A..T...................................T.......................... 382
SiC4-NADP-ME   .H.........V.I...................................................S..V......H.... 385
ZmnonC4-NADP-ME .H..M...A....I...............N..A.........................V..R....H.... 390
SbnonC4-NADP-ME .H..M...A....I....V..........A...................V...V.......H.... 398
Os-NADP-ME     .H..M........I...............A..................SS..V......H.... 385

                          420             440             460             480
                           |               |               |               |
ZmC4-NADP-ME   GAGEAGTGIAELIALEISKQTNAPIEECRKKVWLVDSKGLIVDSRKGSLQPFKKPWAHEHEPLKTLYDAVQSIKPTVLIG 462
SbC4-NADP-ME   ................M....K.....................S..A...........T.............. 462
SiC4-NADP-ME   ...............Q....K.......Q.............S...E..PS..Q......VS..F...... 465
ZmnonC4-NADP-ME ................S............I...........N.V.S...........T.............. 470
SbnonC4-NADP-ME ...............KT......................N...S...........T.............. 478
Os-NADP-ME     ...............K...........L........N...E...A..........VT..L.......... 465

                  *       500             520 *          540             560
                   |               |       |      *        |               |
ZmC4-NADP-ME   TSGVGRTFTKEIIEAMSSFNERPIIFSLSNPTSHSECTAEQAYTWSQGRSIFASGSPFAPVEYEGKTFVPGQSNNAYIFP 542
SbC4-NADP-ME   ...........V..A.I................T..AV..........D............... 542
SiC4-NADP-ME   .....A....VV..A........V..............N.TK.AV......N....D........A...... 545
ZmnonC4-NADP-ME ..........VV...A..................E..N.T...V......D....D..V.......... 550
SbnonC4-NADP-ME ..........VV...A..................E..N.T...A......D....D..I....A...... 558
Os-NADP-ME     .....K....V....A.......V...........E..N.....AV.......D....N..IH........... 545

                  *       580             600             620             640
                   |               |               |               |
ZmC4-NADP-ME   GLGLGLVISGAVRVHEDMLLAASKALADQATQDNFEKGSIFPPFTSIRKISAHIAAAVAAKAYELGLATRLPPPSDLVKY 622
SbC4-NADP-ME   ..................A.......EE..VT........N......Y..............K...A. 622
SiC4-NADP-ME   ......IMA..T.........E......N.E.........S.................A.R...Q. 625
ZmnonC4-NADP-ME F..........I.......E..A...E.D.........N.....R..............R..... 630
SbnonC4-NADP-ME F..........I.......E...G...E.D.........N.....R..............R..... 638
Os-NADP-ME     F...V...........ET.......E.............N.....R...S.............Q.R..E.. 625


ZmC4-NADP-ME   AENCMYTPVYRNYR 636
SbC4-NADP-ME   ..S...S......Q 636
SiC4-NADP-ME   ..T...N.I..... 639
ZmnonC4-NADP-ME .QS.......... 644
SbnonC4-NADP-ME .S........... 652
Os-NADP-ME     ..S........S.. 639
```

Supplementary Fig. 4. Alignment positions of amino acids strictly differentially conserved between plastidic C4- and nonC4-NADP-ME sequences in Poaceae. Asterisks (*) denote the amino acids that are strictly differentially conserved between C4- and nonC4-NADP-ME sequences of maize, sorghum, and rice. The red boxes indicate the amino acids that are strictly differentially conserved in the extended dataset that includes the plastidic C4-NADP-ME of *Setaria italica* (Si). The arrow indicates the transit peptide cleavage site of ZmC4-NADP-ME. Os: *Oryza sativa*.
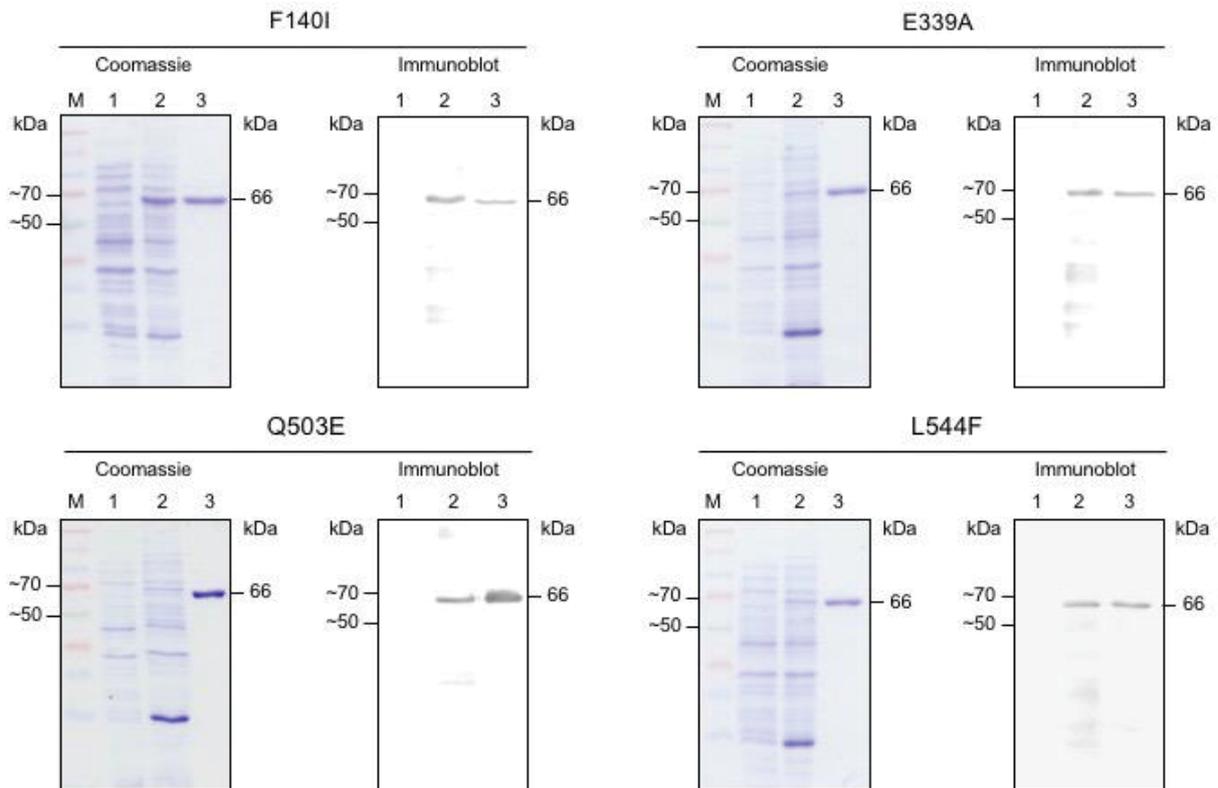
Supplementary Fig. 5. SDS-PAGE stained with Coomassie-Blue and analysed by immunoblot of protein fractions during the isolation of recombinant ZmC4-NADP-ME mutant variants. M, molecular weight markers; 1, 10 µg of non-induced cell culture lysate; 2, 10 µg of cell culture lysate 16 h after induction of protein production; 3, 1.5 µg of affinity-purified ZmC4-NADP-ME mutant variant (1.5 µg for protein staining and 0.5 µg for immunoblot). The size of the recombinant proteins (66 kDa) is indicated on the left. Anti-His-HRP conjugate antibodies were used for the immunoblots.
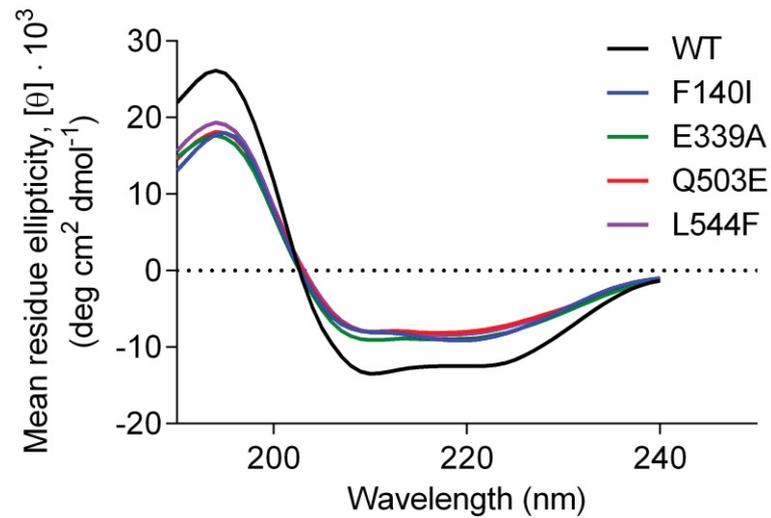
Supplementary Fig. 6. Analysis of the secondary structure of ZmC$_4$-NADP-ME WT and the mutated variants by circular dichroism. Shown are the calculated CD spectra based on the measured scans in the range of 190 to 240 nm at 20 °C with 5 accumulations, each. Data evaluation was performed by the CONTIN-LL algorithm using protein reference set 4 optimised for 190-240 nm (Dichroweb server).

Supplementary Table 1. Protein structure data collection and refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses.

	ZmC$_4$-NADP-ME pdb-code: 5OU5	SbC$_4$-NADP-ME pdb- code: 6C7N
Wavelength [Å]	0.9763 Å	1.5418Å
Resolution range [Å]	48.83 - 2.20 (2.28 - 2.20)	29.28 - 2.00 (2.03 - 2.00)
Space group	I 21 21 21	C 1 2 1
Unit cell [Å; °]	135.96 147.16 261.62 90 90 90	208.10, 64.26, 202.72 90 94 90
Total reflections	781428 (77869)	665459 (31790)
Unique reflections	132076 (12943)	181154 (8879)
Multiplicity	5.9 (6.0)	3.7 (3.6)
Completeness	1.00 (0.99)	1.00 (1.00)
Mean I/sigma(I)	18.7 (3.51)	9.1 (2.1)
Wilson B-factor [Å^2]	36.7	18.2
R-merge	0.063 (0.472)	0.085 (0.585)
R-meas	0.070 (0.515)	0.100 (0.690)
CC1/2	0.999 (0.906)	0.997 (0.733)
CC*	1 (0.975)	1 (0.946)
Reflections used in refinement	132069 (12943)	181074 (17968)
Reflections used for R-free	6717 (638)	1819 (164)
R-work	06 (0.209)	0.180 (0.211)
R-free	0.207 (0.264)	0.210 (0.252)
CC(work)	0.961 (0.812)	0.950 (0.871)
CC(free)	0.950 (0.736)	0.941 (0.764)
Number of non-H atoms	18246	18830
macromolecules	17500	15643
ligands	17	297
Protein residues	2230	2030
RMS(bonds)	0.018	0.010
RMS(angles)	1.83	0.99
Ramachandran favored (%)	98.08	98.20
Ramachandran allowed (%)	1.7	1.8
Ramachandran outliers (%)	0.22	0
Rotamer outliers (%)	3.1	2.0
Clashscore	1.76	1.64
Average B-factor [Å^2]	40.13	33.42
macromolecules	40.13	32.46
ligands	46.00	32.46
solvent	39.83	40.90

Supplementary Table 2. The dimer is formed by strong interactions of residues located in monomers A and B, residues forming the tetramer interface are mainly located in monomers A and C. Given are the interacting amino acids with their numbers, followed by the atom type or identifier. (bb) means "backbone".

Residues involved in dimer-formation (monomer-monomer-interactions)	
Monomer A	Monomer B
Pro202-O (bb)	Tyr98-OH
Asn221-ND2	Arg123-O (bb)
Tyr206-N	Gly124-O (bb)
Ile291-O (bb)	Ala129-N (bb)
Gly196-O (bb)	Lys139-NZ
Glu288-OE1	Lys138-NZ
Gln152-NE2	Gln146-O (bb)
Gln146-O (bb)	Gln152-NE2
Asn159-ND2	Asn159-OD1
Asn159-OD1	Asn159-ND2
Lys139-NZ	Gly196-O (bb)
Leu101-O (bb)	Arg201-NH1
Tyr98-OH	Pro202-O (bb)
Gly124-O (bb)	Tyr206-N (bb)
Arg123-O (bb)	Asn221-ND2
Lys138-NZ	Glu288-OE1
Lys138-NZ	Glu288-OE2
Ala129-N (bb)	Ile291-O (bb)
Residues involved in tetramer-formation	
Monomer A	Monomer C
Val87-N (bb)	Asn634-O (bb)
Pro89-O (bb)	Ala95-N (bb)
Trp90-O (bb)	His106-NE2
Ala91-N (bb)	Ser93-O (bb)
Ala91-O (bb)	Ser93-N (bb)
Ser93-O (bb)	Ala91-N (bb)
Ala95-N (bb)	Pro89-O (bb)
Arg633-NH1	Glu83-O (bb)
Arg633-NH2	Glu83-O (bb)
Asn634-ND2	Glu84-OE1
Asn634-ND2	Glu84-O (bb)
Asn634-O (bb)	Val87-N (bb)

Supplementary Table 3. Hydrogen bonds with and distances to the cofactor NADP in monomer A of SbC4-NADP-ME.

Atom	Residue	Atom	Distance [Å]
ND2	Asn331	O1A	2.75
N	Ala384	O3B	2.92
N	Glu386	O1A	3.08
N	Ala387	O2N	2.82
N	Ser419	O1X	3.19
N	Ser419	O2B	3.26
OG	Ser419	O1X	2.60
NZ	Lys435	O3X	2.73
OG	Ser491	O3D	3.32
N	Asn492	O3D	3.29
OG	Ser535	N7N	2.66
OD1	Asn537	O7N	2.93

5.2 Manuscript II: Posttranslational modification of the NADP-malic enzyme involved in maize C4 photosynthesis fine-tunes the enzymatic activity during the day

Authors: **Anastasiia Bovdilova**, Bruno M. Alexandre, Clarisa E. Alvarez, Inês Matias Luís, Astrid Höppner, David Bickel, Holger Gohlke, Christina Decker, Luitgard Nagel-Steger, Felipe Trajtenberg, Alejandro Buschiazzo, Saleh Alseekh, Alisdair R. Fernie, Maria F. Drincovich, Isabel A. Abreu, Veronica G. Maurino

This manuscript was submitted to The Plant Cell in the year 2018.

Contribution of Anastasiia Bovdilova:

In vitro mutagenesis and cloning of expression constructs

Isolation of recombinant proteins

Measurement of kinetic parameters

Circular dichroism measurements

Measurement of malate and NADP-ME activities in maize leaves extracts

Figure design and drafting of the manuscript

Posttranslational modification of the NADP-malic enzyme involved in maize C[4] photosynthesis fine-tunes the enzymatic activity during the day

Anastasiia Bovdilova[1], Bruno M. Alexandre[2,3], Clarisa E. Alvarez[4], Inês Matias Luís[2], Astrid Höppner[5], David Bickel[6], Holger Gohlke[6], Christina Decker[7], Luitgard Nagel-Steger[7], Felipe Trajtenberg[8], Alejandro Buschiazzo[8], Saleh Alseekh[9,10], Alisdair R. Fernie[9,10], Maria F. Drincovich[4], Isabel A. Abreu[2,3], Veronica G. Maurino[1],*

[1]Institute of Developmental and Molecular Biology of Plants, Plant Molecular Physiology and Biotechnology Group, Heinrich Heine University, Universitätsstraße 1, and Cluster of Excellence on Plant Sciences (CEPLAS), 40225 Düsseldorf, Germany.

[2]Instituto de Biologia Experimental e Tecnológica, Avenida da República, Quinta do Marquês, 2780-157 Oeiras, Portugal.

[3]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Avenida da República, 2780-157 Oeiras, Portugal.

[4]Centro de Estudios Fotosinteticos y Bioquimicos (CEFOBI-CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, University of Rosario, Suipacha 570, 2000 Rosario, Argentina.

[5]Center for Structural Studies (CSS), Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[6]Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[7]Institut für Physikalische Biologie, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany and Institute of Complex Systems, Structural Biochemistry (ICS-6), Forschungszentrum Jülich, 52425 Jülich, Germany.

[8]Laboratory of Molecular and Structural Microbiolgy, Institute Pasteur Montevideo, Mataojo 2020, 11400 Montevideo, Uruguay.

[9]Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.

[10]Center for Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria.

* Corresponding Author: veronica.maurino@uni-duesseldorf.de

Short Title: *In vivo* phosphorylation of maize C$_4$-NADP-ME

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Veronica G. Maurino (veronica.maurino@uni-duesseldorf.de).

ABSTRACT

Evolution of the C$_4$ photosynthetic pathway involved recruitment of housekeeping proteins through gene duplication and their further neofunctionalization. NADP-malic enzyme (ME), the most widespread C$_4$ decarboxylase, has increased its catalytic efficiency and acquired regulatory properties that allowed it to participate in the C$_4$ pathway. Here, we show that regulation of maize C$_4$-NADP-ME activity is much more complex than previously reported. Mass spectrometry analysis identified phosphorylation of the serine 419 (Ser419) of C$_4$-NADP-ME in protein extracts of maize leaves. The phosphorylation event increases after the light turns on, with a peak at ZT2. Phosphorylation of ZmC$_4$-NADP-ME drastically decreases its activity as shown by the low residual activity of the recombinant phosphomimetic mutant. Analysis of the crystal structure of C$_4$-NADP-ME in complex with the cofactor indicated that Ser419 is involved in the binding of NADP at the active site. Modeling and molecular dynamics simulations support less favorable binding of the cofactor NADP in the phosphomimetic mutant. We propose that phosphorylation of ZmC$_4$-NADP-ME at Ser419 during the first hours in the light is a cellular mechanism to fine-tune the enzymatic activity to coordinate the carbon concentration mechanism with the CO$_2$ fixation rate, most probably to avoid CO$_2$ leakiness from bundle sheath cells.

INTRODUCTION

The world's most productive crops perform C$_4$ photosynthesis. The ancestors of these plants evolved a biochemical pump to concentrate CO$_2$ at the site of Rubisco. Leading to lower photorespiratory fluxes and greater photosynthetic efficiency (Furbank and Hatch, 1987). Compared with the ancestral C$_3$ photosynthesis, the C$_4$ pathway allows increased plant productivity in warm habitats due to a more efficient use of nitrogen and water.

The most agronomically important C_4 plants maize, sorghum, and sugar cane belong to the NADP-ME subtype. In these plants CO_2 is initially fixed in the mesophyll cells by phosphoenolpyruvate carboxylase (PEPC), leading to the formation of oxaloacetate. Oxaloacetate is then predominantly reduced to malate and transported into the bundle sheath cells (BSC), where CO_2 is released by NADP-malic enzyme (NADP-ME) for its refixation through Rubisco (Hatch, 1987). By this process, CO_2 concentration in the BSC rises to levels up to 1,500 µL L^{-1} (Furbank and Hatch, 1987;Sage et al., 2012).

The transition from C_3 to C_4 metabolism involved complex alterations to leaf anatomy and biochemistry. A crucial step in the evolution of the C_4 photosynthetic pathway was the recruitment of enzymes through gene duplication of housekeeping isoforms and subsequent neofunctionalization. In the course of its evolution, the C4-specific NADP-ME isoform increased its catalytic efficiency (Saigo et al., 2013) and acquired regulatory properties that optimized its efficiency in the C_4 pathway. During the night, when the delivery of malate from the PEPC reaction is stopped, C4-NADP-ME activity is inhibited by malate (Saigo et al., 2004;Saigo et al., 2013). This property exclusive to the C4 isoform minimizes the malate consumption to avoid extreme carbon loss during the night period, which would otherwise result in carbon starvation (Fahnenstich et al., 2007;Zell et al., 2010). At the pH of the stroma in darkness, a proportion of C4-NADP-ME likely loses its quaternary structure and adopts a lower oligomerization state that is less active or even inactive (Iglesias and Andreo, 1990). This likely represents a second level of regulation that ensures that C4-NADP-ME is less active during the night period.

The enzymatic activity of other C_4 proteins in known to be regulated through post-translational modifications (PTMs). PTMs provide mechanisms for rapid, reversible control of C_4 protein function especially during the day-night transitions and in response to light intensity (Walker et al., 2002;Chen et al., 2014). The activity of C4-PEPC is regulated through phosphorylation of a serine residue located to the N-terminal region. C4-PEPC is mainly phosphorylated in the light, where it has a higher catalytic activity and is less sensitive to the allosteric inhibition by malate than in darkness (Jiao and Chollet, 1988; 1989; 1990). C4-pyruvate orthophosphate dikinase (C4-PPDK) activity, in contrast, is not controlled by the light/dark transition per se but by the light intensity; phosphorylation decreases (and activity increases) with mounting light intensity in the morning and increases with waning light intensity in the evening evening (Chen et al., 2014). C4-PPDK is regulated by reversible phosphorylation of a threonine residue of the active site (Burnell and Hatch, 1985;Chastain et al., 2000). In Guinea grass (*Panicum*

maximum) and maize, C_4-PEP carboxykinase (C_4-PEPCK) is phosphorylated in the dark when the enzyme is less active, and dephosphorylated under illumination (Walker and Leegood, 1996;Walker et al., 2002;Chao et al., 2014). Changes in the phosphorylation state of PEPCK lead to changes in its sensitivity to regulation by adenylates, which likely activate the enzyme in the light (Walker et al., 2002). Recently, it was shown that phosphorylation of Ser55 of maize C_4-PEPCK is also dependent on the light regime (Chao et al., 2014).

Little is known about the participation of PTMs in the regulation of the C_4-NADP-ME isoform *in vivo*. Until now, only the effects of redox modulation on the recombinant ZmC_4-NADP-ME activity were reported (Alvarez et al., 2012). Oxidation of ZmC_4-NADP-ME decreases the catalytic activity and increases the affinity for malate and the cofactor; in combination, these changes produce no significant changes of the catalytic efficiency. The oxidation of Cys192, Cys246, Cys270 and Cys410 may contribute to the changes observed *in vitro* (Alvarez et al., 2012).

Here, we show that ZmC_4-NADP-ME is phosphorylated *in vivo* at Ser419. The phosphorylated fraction increases after the light is turned on, with a peak at 2 h after the onset of light. We found that the recombinant phosphomimetic mutant S419E presents a low residual activity at stromal pH during the day, indicating that phosphorylation of ZmC_4-NADP-ME inactivates the protein. We solved the crystal structure of *Sorghum bicolor* C_4-NADP-ME in complex with NADP and found that Ser419 is involved in the binding of NADP at the active site. Molecular modeling and dynamics simulations of ZmC_4-NADP-ME complexes revealed that NADP binding in the S419E mutant is energetically less favorable and that the cofactor has higher mobility in the mutant´s binding site. We postulate that ZmC_4-NADP-ME phosphorylation at Ser419 is a mechanism to fine-tune the enzymatic activity according to the operation of the C_4 pathway, most probably to coordinate malate decarboxylation with carbon fixation in BSC.

RESULTS

Identification of a phosphorylation site in ZmC₄-NADP-ME

Mass spectrometric identification of phosphorylation sites faces several challenges. Phosphopeptides are often of low abundance in complex protein extracts such as cell lysates obtained from maize leaf macerates. Also, phosphopeptides often exhibit low MS signals and inadequate fragmentation patterns. To overcome these issues and to be able to identify phosphorylated peptides in ZmC4-NADP-ME, we loaded maize protein extracts onto 4-12% NuPAGE gels and electrophoresed them for a short time. Upon in-gel digestion, samples were subjected to a phosphopeptide enrichment step and analyzed by LC-MS using an AB Sciex TripleTOF 6600 mass spectrometer. The acquired data was analyzed with Protein Pilot 5.0 and PeakView 2.2. This approach was successful in unambiguously detecting the phosphorylation of ZmC4-NADP-ME at a single amino acid residue, Ser419. Figure 1 shows a tandem MS spectrum of the obtained peptide acVWLVDpSK. The quality of the MS/MS data for this peptide enabled the correct assignment of the phosphorylated residue. In fact, with the exception for the y_1-ion, the entire y-ion series of the peptide was assigned, and the loss of the phosphate group (H_3PO_4), corresponding to a mass-to-charge ratio of 98, was detected for the whole y-ion series.

The MS analysis did not identify any phosphorylation event in the other maize NADP-ME isoforms, as the Val414 from the identified phosphorylated peptide is unique to the ZmC4NADP-ME (Alvarez et al., 2013). We conclude that a functional role for this phosphorylation site most probably represents a C4-trait. We next compared the protein sequences of plastidic photosynthetic and non-photosynthetic NADP-ME isoforms within the Panicoideae subfamily of the Poaceae. This sequence alignment indicated that the serine corresponding to amino acid position 419 in the C4-NADP-ME protein sequence is conserved among all NADP-ME sequences analyzed (Supplemental Figure 1).

Production of the phosphomimetic mutant and characterization of the biochemical properties

In order to analyse the effect of phosphorylation of Ser419 on the biochemical properties of ZmC4-NADP-ME, we generated the phosphomimetic mutant S419E by changing the Ser419 to Glu by site-directed mutagenesis. Phosphorylation adds negative charge to amino acids, and thus the negatively charged amino acid Glu can be used to mimic the phosphorylated state of a protein (Dissmeyer and Schnittger, 2011). In order to assess the importance of the amino acid

position for the enzymatic activity, we also created the S419A variant by changing the Ser419 for Ala, a small non-polar amino acid.

ZmC4-NADP-ME wild-type (WT), S419E, and S419A were heterologously expressed and purified to homogeneity by affinity chromatography. The recombinant proteins, showing in all cases the expected molecular masses (Supplemental Figure 2), were subjected to a comparative analysis of their enzymatic properties. Analysis of the dependence of the enzymatic activity on the pH of the reaction media indicated that while the recombinant WT ZmC4-NADP-ME has a pH optimum of 8.0, the S419E variant presents a shift of the pH optimum towards more acidic pHs between 6.5 and 7.0 (Figure 2A). The S419A variant showed a pH optimum of 7.0 - 7.5 (Figure 2A). At the physiologically relevant pH 8.0 of the photosynthetically active stroma, the S419E variant presents only 8.7% of the WT activity, while the S419A variant retains a 72.3% of the WT activity (Figure 2A).

Analysis of the affinity to NADP indicated that the exchange of the Ser residue for Glu in the S419E variant dramatically decreases the affinity of the mutant for the cofactor, as saturation was not observed even at NADP concentrations as high as 5 mM. The S419A variant presented a 13-fold lower affinity (K_m 159.5 μM) for the cofactor in comparison to the WT (K_m 12.3 μM) (Figure 2B), indicating that S419 is essential for the normal operation of the enzyme.

The affinity to the substrate malate at pH 8.0 was measured using 0.5 mM NADP, which was saturating for S419A, but not for S419E. In comparison to the WT (K_m= 0.19 mM), the apparent affinity for malate decreased by approximately 14-fold in the S419E variant (K_m =2.68 mM) and only 2-fold in the S419A variant (K_m = 0.47 mM). The turnover number (k_{cat}) was measured at pH 8.0, at saturating levels of malate, and at 0.5 mM NADP. Under these conditions, the S419E variant displayed a 6-fold lower catalytic activity (k_{cat} 2.1 s^{-1}) with respect to the WT (k_{cat} 12.2 s^{-1}) (Figure 2B). In contrast, the k_{cat} of the S419A variant (k_{cat} 8.3 s^{-1}) was in the same range as that of the WT (Figure 2B).

Structural organization of the mutant variants

As the phosphomimetic mutant showed dramatic changes of its kinetic properties in comparison to the WT enzyme, we analysed if these could be the result of alterations in the structural organisation of the engineered protein. We used circular dichroism (CD) to explore the secondary structure organization and analytical ultracentrifugation (AUC) to explore the quaternary composition.

CD measurements at both pH 7.0 and 8.0 indicated no significant changes in the secondary structure organization of the mutated variants in comparison to the WT (Figure 3A and B). The data obtained indicated that ZmC4-NADP-ME WT and the S419E and S419A variants presented similar fractions of different types of secondary structures depending on the pH. At pH 8.0, all enzymes possess almost twice the amount of helical structures observed at pH 7.0, an approximately 1.5-fold lower amount of β-strand structures, an almost unchanged amount of turns, and an only slightly decreased amount of unordered secondary structures (Supplemental Table 1).

AUC measurements conducted at both pH 7.0 and 8.0 (Figure 3C and D) confirmed a stable tetrameric state of WT ZmC4-NADP-ME at pH 8.0, with a sedimentation coefficient of 11.7 S (Figure 3D). Both mutated variants also assemble as tetramers at pH 8.0, showing sedimentation coefficients of 11.6 S and 11.1 S (Figure 3D). The sedimentation coefficient distribution of all proteins at pH 7.0 showed two peaks, with sedimentation coefficients of approximately 11.5 S and 5.4 S (Figure 3C), indicating that at this pH the enzymes exist as a mixture of monomers/dimers and tetramers with predominance of the latter. The results also indicate that the generated mutant variants have no altered quaternary structure at any pH in comparison to the WT. It can be speculated that the changes in amount of helical and turn components observed in the CD measurements at the different pHs might be related to the occurrence of a dimeric fraction at pH 7.0, as these changes may contribute to the alterations in the interface contacts between the monomers in the tetramer.

Taken together, the results of CD and AUC analyses indicate that the observed changes in the kinetic properties of the mutant variants are not due to changes in the structural organization of the proteins with respect to the WT at each pH analysed.

Diurnal profiles of ZmC4-NADP-ME phosphorylation state and NADP-ME activity in maize leaves

We quantified ZmC4-NADP-ME total protein amount in leaves of maize at different time points during a day using SWATH-MS (sequential window acquisition of all theoretical fragment ion spectra-mass spectrometry). We found that during the day, there are no major fluctuations of the protein levels, apart from a small decrease at ZT4 (Figure 4A). Following the light-to-dark transition, there is an increase in total ZmC4-NADP-ME protein level (Figure 4A). Despite a decreasing trend over the dark perio from ZT16.5 to ZT23.5, ZmC4-NADP-ME

protein levels (as measured by SWATH-MS) are higher during the whole dark period when compared to the light period.

Quantification of phosphorylation in Ser419 throughout the day indicated a significant increase of phosphorylation events after the light is turned on, with a peak at ZT2 that accounts for about 13% of the total ZmC4-NADP-ME protein. This peak is followed by a substantial decrease to phosphorylation levels as low as 3% at ZT4 in the light. These low levels are maintained for the remainder of the day and also during the night (Figure 4B).

Measurements of total NADP-ME activity in extracts of maize leaves indicated a slight increment of the activity between ZT2 and ZT4 in the light, followed by a constant decline during the rest of the light period (Figure 4C); however, these putative changes are not statistically significant over the course of the day. The activity measured in total leaf extracts is the sum of the activity of all maize NADP-ME isoforms. As the C4 isoform is the most highly expressed isoform in photosynthetic tissues and the other isoforms likely have a constant, lower activity over the diurnal cycle (Detarsio et al., 2008), we conclude that the tendencies observed in the profile of measured NADP-ME activity represents the activity of the C4 isoform.

Concentration of malate and related metabolites over the diurnal cycle in maize leaves

We next investigated the profile of malate concentration in whole maize leaf extracts during a diurnal cycle (Figure 4D). Similar high malate concentrations were observed during the light period and at ZT16.5 (530 to 670 nmol g^{-1} fresh weight). These values decreased, although not statistically significantly, during the dark period, reaching a value of approx. 340 nmol g^{-1} fresh weight by the end of the night (Figure 4D). This profile of malate concentrations in maize leaves is different from that described in some C3 plants such as *Arabidopsis*, where malate functions as a transient carbon storage molecule. In these C3 plants, malate concentrations increase continuously during the day and decrease during the night (Fahnenstich et al., 2007;Zell et al., 2010).

In maize leaves, malate not only participates in C4 metabolism but also has a major involvement in basal mitochondrial respiration as part of the tricarboxylic acid cycle (TCA). Thus, we measured the relative levels of other primary metabolites to analyze if the constantly high levels of malate found during the day influence the levels of other primary metabolites during a diurnal period and this could be related to the observed regulation of C4-NADP-ME (Supplemental Table 2). The analysis of the relative levels at different time points during the

day with respect to those found by the end of the night (ZT23.5) of each metabolite allowed to separate them into two clusters (A and B; Figure 5). Cluster A groups several metabolites that, relative to the levels found at the end of the night, present low levels during most of the light period, especially towards the end of the day and/or the beginning of the night. Cluster B, which contains the majority of the metabolites analyzed, groups metabolites that present high relative levels by the end of the day and/or beginning of the night with respect to those found by the end of the night. In cluster B, the metabolites could be further grouped into sub-clusters with specific patterns (Figure 5).

We were able to determine the levels of almost all intermediates of the TCA cycle except of oxaloacetate and isocitrate; we could however determine the levels of *cis*-aconitate, the intermediate in the isomerization of citrate to isocitrate. All the TCA organic acids are found in cluster B, indicating that they are present at higher levels during the day or at the beginning of night than at the end of the night. We propose that this likely reflects a higher rate of respiration at the beginning of the day and thus the partial inactivation of C4-NADP-ME would be a means to reduce carbon loss from the BSCs due to CO_2 leakiness (the proportion of carbon fixed by PEPC which subsequently leaks out of the BSC).

The diurnal fluctuations of fumarate levels are similar to those described in C3 plants (Fahnenstich et al., 2007). Fumarate increases during the day and decreases during the night, while as quoted before (Figure 4D), malate maintains a quite constant level during the day and decreases by the end of the night. This indicates that the pool of malate participating in C4 metabolism is likely separate from that involved in the TCA cycle. This observation agrees with the fact that very little carbon leaks from the large pools of metabolites of the C4 pathway into respiratory metabolism in maize (Arrivault et al., 2017).

The oxaloacetate-derived amino acids aspartate group in cluster A, with lower levels during most of the day and the beginning of the night compared to the end of the night. In leaves of *Arabidopsis*, aspartate displays much higher levels during the day than by the end of the night (Fahnenstich et al., 2007). This difference in accumulation patterns during a diurnal cycle observed between C3 and C4 leaves may be related to the proposed use of aspartate, in parallel to malate, to shuttle CO_2 to the bundle sheath cells in the leaves of maize (Pick et al., 2011;Arrivault et al., 2017). Under this hypothesis, one would expect that aspartate relative levels would be comparable to those of malate during the photosynthetic period. Indeed, the levels of aspartate increase during the day and decrease at the beginning of the night (Supplemental Table 2).

Amino acids derived from pyruvate, a C_4 cycle and TCA cycle intermediate, show a pattern similar to malate, with higher levels during the day and at the beginning of the night period than at the end of the night. While the patterns of valine and leucine in maize leaves are comparable to those observed for *Arabidopsis*, that of alanine differs considerably. Alanine levels are much higher during the day in the leaves of C_4 than in C_3 leaves, indicating that pyruvate produced during the decarboxylation of malate through NADP-ME might be used for the synthesis of alanine. This is in accordance with a recently revised model of the C_4 cycle, which posits that alanine would be produced from pyruvate through an alanine amino transferase and shuttled to the mesophyll cells for the regeneration of the carbon acceptor (Pick et al., 2011).

The pattern of accumulation of the aromatic amino acids tryptophan, phenylalanine and tyrosine, which are formed from PEP via the shikimate pathway, indicates that in maize, PEP is preferentially channeled into the synthesis of the aromatic amino acids when photosynthesis is inactive, while during the day it is involved in C_4 metabolism. These aromatic amino acids cluster together presenting higher relative levels at the beginning of the night and lower levels during the day. These patterns contrast with those described in leaves of *Arabidopsis*, where the lowest levels are found at the end of the night (Fahnenstich et al., 2007).

Localization of Ser419 in the C_4-NADP-ME structure

To aid in identifying the influence of Ser419 phosphorylation on the biochemical changes observed in C_4-NADP-ME, we solved the crystal structure of SbC_4-NADP-ME as a complex with NADP at 2.0 Å resolution (pdb-code 6C7N) and analyzed the cofactor binding site. Sorghum belongs to the same C_4 linage as maize and the C_4-NADP-ME of these species share 85% amino acid identity (Emms et al., 2016). Data collection and refinement statistics are summarized in Supplemental Table 3. The SbC_4-NADP-ME contains a homo-tetramer in the asymmetric unit, whereas only two monomers (chains A and B) are traceable unambiguously in the electron density for the entire monomer (Figure 6A). The sections from aa 360-530 in chains C and D show either no or very poor electron density, indicating flexible parts in the crystal. The cofactor NADP is bound in both monomers A and B and well defined by the electron density. We identified the amino acids involved in cofactor binding and measured their hydrogen bonding distances (Figure 6B and C, Supplemental Table 4). These results show that Ser419 is directly involved in NADP binding.

Using the structural information, we also found that Ser419 is located on the surface of the enzyme. It might, therefore, be spatially available for the interaction with a kinase/phosphatase (Supplemental Figure 3). Analysis of known kinase/phosphatase binding motifs associated with the identified phosphorylation site in ZmC4-NADP-ME indicated the presence of a β-adrenergic receptor kinase (BARK) substrate motif (Amanchy et al., 2007), DSKGL (Supplemental Figure 1).

NADP binding is less stable in the S419E variant than in wild type ZmC4-NADP-ME

In order to assess the stability of the binding mode of NADP in WT ZmC4-NADP-ME and the S419E variant, we performed molecular dynamic (MD) simulations of 300 ns on the respective complexes. The C_α-RMSD for ZmC4-NADP-ME, a measure for internal solute motion with respect to the starting structure (excluding highly flexible loop regions in the N- and C-termini), revealed values between 2 Å and mostly < 4 Å (Figure 7A). In contrast, the RMSD of NADP computed after superimposing only the protein C_α atoms only displays a marked difference between the WT and variant complexes: while NADP bound to WT ZmC4-NADP-ME remains close to the initial binding mode throughout the simulation time (mean RMSD < 3 Å), NADP in S419E partially unbinds after ~50 ns of simulation time and then remains in a configuration with an RMSD of ~5 Å thereafter (Figure 7B). These findings suggest that binding of NADP to S419E is less favorable than to WT ZmC4-NADP-ME.

In order to corroborate this finding, effective binding energies for NADP to either system were computed following the MM-GBSA approach (Kuhn and Kollman, 2000;Hou et al., 2011). The time series of the energies (see Figure 7C) reflect the ligand's motional behavior in that increases (decreases) in the effective energy of one system are largely paralleled by respective increases (decreases) in the ligand's RMSD (Figure 7B) and loss (formation) of non-covalent interactions found in the starting structure (Figure 7D).

The effective binding energy for NADP binding to WT ZmC4-NADP-ME is by 17.1 ± 0.3 kcal mol^{-1} more favorable than that of NADP in the mutant S419E (NADP in WT: -11.4 ± 0.2 kcal mol^{-1}; NADP in S419E: -28.5 ± 0.2 kcal mol^{-1}). The difference in the effective binding energies is significant according to a Student's t-test ($p < 0.001$). Note, though, that configurational entropy differences were neglected here, such that the difference in the effective binding energies may be smaller due to enthalpy-entropy compensation effects. These results indicate a lower binding affinity of NADP to S419E than to WT, which agrees with the observation that

no NADP saturation could be reached during biochemical characterization of S419E (Figure 2). Furthermore, the higher mobility of the cofactor in S419E's binding site may explain the ~ 6-fold lower rate constant for the S419E variant compared to the WT, as it makes reaching a reactive configuration with malate less likely.

DISCUSSION

ZmC$_4$-NADP-ME is phosphorylated *in vivo* at a serine that is directly involved in NADP binding

Through mass spectrometry, we identified a single phosphorylation event of ZmC$_4$-NADP-ME at serine 419 in maize leaves. Phosphorylation levels increase after the light is turned on, with a peak at ZT2. Through mutational analysis and biochemical characterization of ZmC$_4$-NADP-ME WT and enzyme variants, we demonstrate that the phosphomimetic mutant S419E has highly reduced affinity to the cofactor NADP as well as lower catalytic activity compared to the non-phosphorylated enzyme. S419E presented a low residual activity at the stromal pH in the light, indicating that phosphorylation of ZmC$_4$-NADP-ME drastically decreases its activity. After solving the crystal structure of SbC$_4$-NADP-ME as a complex with NADP, we found that Ser419 is located on the surface of the enzyme, which makes it spatially available for the interaction with a kinase/phosphatase. Moreover, analysis of the crystal structure indicates that Ser419 is directly involved in NADP binding. Molecular dynamics simulations and free energy calculations revealed that the effective binding energy for NADP binding is less favorable to binding in the phosphomimetic mutant S419E. Additionally, the higher mobility of the cofactor in the variant's binding site leads to binding positions of NADP that are not in a suitable reactive configuration with the substrate. This explains the highly reduced affinity of S419E for NADP and indicates that phosphorylation of Ser419 modulates the activity of ZmC$_4$-NADP-ME by creating unfavorable binding conditions for its cofactor.

Furthermore, we observed conservation of the Ser419, as well as of a β-adrenergic receptor kinase (BARK) substrate motif (DSKGL), which includes the Ser419, among plastidic photosynthetic and non-photosynthetic NADP-ME isoforms within the Poaceae (Supplemental Figure 1). This pattern of sequences conservation and the fact that only the C$_4$-NADP-ME isoform is phosphorylated suggests that the use of reversible phosphorylation of Ser419 to control the activity of NADP-ME appear with the evolution of the C$_4$ pathway and should involve a BSC chloroplastic kinase, which might have been co-opted for this specific task.

Phosphorylation of ZmC4-NADP-ME at Ser419 is a mechanism to fine-tune the enzymatic activity

The operation of the C4 pathway requires a deep synchronization between mesophyll and BSCs during a day-night cycle (Bailey et al., 2007). To efficiently concentrate CO_2 in BSC chloroplasts, coordination of C4-NADP-ME decarboxylation and Rubisco carboxylation rates is imperative as this will prevent loss of the CO_2 released by C4-NADP-ME. During the night, when the photosynthetic pathway is inactive, C4-NADP-ME activity is regulated through two processes: on the one hand, the enzyme is allosterically inhibited by malate, on the other hand, it partially loses its active quaternary oligomerization state (Iglesias and Andreo, 1990;Saigo et al., 2004;Saigo et al., 2013).

Production of malate in mesophyll cells and its decarboxylation in BSCs must also be coordinated to accommodate the large and rapid flux changes that occur during the day. We show that malate concentrations in total maize leaf extracts decrease during the night and increase after the onset of light, quickly reaching values that are maintained during the whole light period. The rapid increment of malate levels at the beginning of the day is likely mostly due to the activity of the C4 concentrating mechanism.

Apart from the contribution of malate decarboxylation, respiration of CO_2 in BSC also contributes to the carbon concentrating mechanism (Bellasio and Griffiths, 2014). It is possible that, at ZT2, respiration in BSC is high enough to make an important contribution to CO_2 fixation by Rubisco. In such conditions, decarboxylation of malate by a fully active C4-NADP-ME could lead to high CO_2 levels in BSC resulting in increased wasteful back leakage of CO_2 to mesophyll cells (Kromdijk et al., 2014). The naturally evolved efficiency of C4 photosynthesis requires the tight regulation of CO_2 supply to Rubisco within the BSCs in order to minimize leakiness and associated energy costs (Furbank et al., 1990). As a consequence, a decrease in C4-NADP-ME activity would avoid extra waste of ATP for the operation of the C4 carbon concentration mechanism. The cooperation of the PEPC-driven carbon concentration and respiration-driven mechanisms require plasticity, as the extent in which each of these mechanisms is involved may depend on changes of environmental conditions such as photosynthetically active radiation, shade by canopy, and water availability. Phosphorylation of ZmC4-NADP-ME at Ser419 during the first hours in the light might be a mechanism to fine-

tune the enzymatic activity, helping to adjust the use of malate in the BSC to avoid CO_2 leakiness and energy waste.

Regulation of C4-NADP-ME activity via phosphorylation may also be important in another context. Aspartate movement to maize BSCs carries approximately 4% of the CO_2 for the C3 cycle. Approximately 40% of this C4-acid are decarboxylated by PEPCK (Arrivault et al., 2017); the remainder aspartate is converted to malate and decarboxylated through NAD(P)-ME. There exists a flexible partitioning of C4 decarboxylation activity between NADP-ME and PEP-CK in response to environmental conditions in maize (Furbank, 2011;Sharwood et al., 2014). In this context, we speculate that phosphorylation of ZmC4-NADP-ME at Ser419 might be a mechanism to rapidly decrease the activity of C4-NADP-ME and with this to switch the use of malate to aspartate according to the metabolic demands imposed by a constantly changing environment.

A currently highly popular goal within the plant scientific community is the improvement of photosynthesis by introducing the C4 concentrating mechanism into C3 plants. Our results are of great importance in this context, as they describe a novel regulation of a C4 enzyme critical for an efficient C4 cycle. Further work should be now directed towards elucidating the environmental conditions in which this regulatory mechanism is required, and if it effectively contributes to the coordination of the C4 and C3 cycles and hence to the metabolic flexibility of C4 plants in response to fluctuating environments.

METHODS

Plant growth conditions

For the mass spectrometry analysis, *Zea mays* (B73) seeds were germinated for 48h in the dark at 28ºC and then transferred to a soil:turf (1:1) mixture. Plants were grown under the following conditions: light/dark cycles of 16/8 h at 28ºC/26ºC; light intensity of 450 µmol m^{-2} s^{-1}. For metabolite analysis and determination of total NADP-ME enzyme activity, plants were grown in Floraton I soil in a Conviron® E15 plant growth chamber under following conditions: light/dark cycles of 16/8 h at 25ºC/20ºC; light intensity of 700 µmol m^{-2} s^{-1}. Maize leaves were collected at ZT (Zeitgeber time) 0.5, 2, 4, 8, 15.5, 16.5, 20 and 23.5; where ZT0 = light on and ZT16 = light off. One-third (from leaf tip) of maize leaves (3rd and 4th leaf) from 12-day-old seedlings were harvested, immediately frozen in liquid nitrogen and stored at -80 °C until further analysis.

Sample preparation for mass spectrometry analysis

Frozen maize leaves were ground with mortar and pestle in liquid nitrogen until a fine powder was obtained. Protein extraction was performed directly in Lysis Buffer as previously described (Luis et al., 2016). For the total protein analysis, protein extracts were loaded onto 4-12% bis-tris polyacrylamide gels (NuPAGE, ThermoFisher Scientific, Waltham, MA USA) and electrophoresed for a short time (10-15 min). The whole bands containing the entire protein extract (2-3 cm long) were excised and in-gel digested with trypsin followed by peptide desalt with C18 tips (Thermo Scientific, Rockford IL USA). For the phosphopeptide analysis of NADP-ME, protein extracts were loaded onto the same kind of gels (4-12% bis-tris polyacrylamide gels, NuPAGE, ThermoFisher Scientific, Waltham, MA USA), electrophoresed at a constant voltage of 25V for 10 min and then for 150V until the Bromophenol Blue reached the bottom of the gel. Then, the gel fractions containing the NADP-ME (ranging from just below the 63 kDa to just after the 75 kDa) were excised, in-gel digested with trypsin and further lyophilized to dryness. Vacuum dried peptides were first desalted using C18 tips (Thermo Scientific, Rockford IL USA) before being subjected to phosphopeptide enrichment using TiO_2 beads (TiO_2 Mag Sepharose, GE Healthcare, Uppsala, Sweden) according to manufacturer's instructions.

LC-MS using an AB Sciex TripleTOF 6600

Samples were analyzed using a TripleTOF 6600 mass spectrometer (AB Sciex, Framingham, MA USA), coupled to a nanoLC Eksigent 425 system (AB Sciex, Framingham, MA USA). RP-HPLC was performed in a trap and elution configuration using a nano cHiPLC trap column (Eksigent ChromXP C18-CL, 3 μm particle size, 120 Å pore size, 0.5 mm x 200 μm I.D., AB Sciex, Framingham, MA USA) and an analytical column (Eksigent ChromXP C18-CL, 3 μm particle size, 120 Å pore size, 15 cm x 75 μm I.D., AB Sciex, Framingham, MA USA). Samples were loaded into the trap column at a flow rate of 2 μL min^{-1} for 10 min using 100% solvent A (0.1% formic acid in water) and eluted at a flow rate of 300nL/min using a stepwise gradient: 0-1min, 4.5% B (0.1% formic acid in acetonitrile); 1-91min, 29.7% B; 91-93, 79.2% B; 93-108min, 79.2% B; 108-110min, 4.5% B; 110-127min, 4.5% B. Samples were run in information-dependent acquisition (IDA) mode to perform peptide and protein identification in order to generate a spectral library for the SWATH quantification. This spectral library was

created by combining all the IDA.wiff files in unison using ProteinPilot 5.0 (AB Sciex). Database search was performed using the Paragon algorithm, which is embedded in ProteinPilot software, v. 5.0 (AB Sciex, Framingham, MA USA). A Paragon search method was created with the following settings: sample type, identification; cys alkylation, acrylamide; digestion, trypsin; instrument, tripleTOF 6600; special factors, phosphorylation emphasis and gel-based ID; species, none; ID focus, biological modifications; search effort, thorough; detected protein threshold, >0.05. The searches were conducted using a UniProt database containing all reviewed protein sequences for the species *Zea mays*. A false discovery rate threshold was set to below 1%.

As for the SWATH-MS analysis, four replicates from each one of the seven time points collected were analyzed by LC-MS using the setup described for the IDA runs. SWATH-MS data were acquired with a standard SWATH acquisition method, using a set of 32 overlapping SWATH windows covering the precursor mass range of 400–1200 m/z. For each SWATH window, the Q1 transmission width was 26 m/z (containing 1 m/z for the window overlap). At the beginning of each cycle, a 50 ms survey scan (350-1250 m/z) was acquired for instrument calibration, and the subsequent SWATH windows were collected from 100 to 1800 m/z for 60 ms, resulting in a cycle time of 2.02 s. Data processing was performed using SWATH processing plug-in for PeakView 2.2 (AB Sciex, Framingham, MA USA). Briefly, peptides were confirmed by finding and scoring peak groups, which are a set of fragment ions for the peptide. Target fragment ions, up to 5, were automatically selected as previously described (Lambert et al., 2013) criteria: (i) fragment ions for a selected peptide were ranked according to ion intensity; (ii) ions higher in m/z than the y4 fragment ion for each selected peptide were ranked highest; (iii) ions within the SWATH isolation window were excluded from selection; (iv) if insufficient target ions were found, ions lower than y4 but outside of the SWATH window were chosen; (v) if there were still insufficient ions, then fragment ions from within the SWATH window region were chosen. Peak group confidence threshold was determined based on an FDR analysis using the target-decoy approach, and 1% extraction FDR threshold was used for all the analyses. Peptides that met the 1% FDR threshold in all the four replicates were retained, and the peak areas of the target fragment ions of those peptides were extracted across the experiments using an extracted-ion chromatogram window of 12 min. ZmC_4-NADP-ME protein levels were estimated by summing the areas of all the transitions from all the peptides for a given protein and normalized to the total area (Collins et al., 2013). As for the ZmC_4-NADP-ME phosphopeptide (acVWLVDpSK) levels, they were estimated by summing

the areas of the transitions obtained for this phosphopeptide and normalized to the sum of the area of all the transitions from all the peptides within the same region (that include Ser419) for each of the time points.

Site-directed mutagenesis of ZmC$_4$-NADP-ME

As template for the PCR reaction (first step of the site-directed mutagenesis), pET32*::*WT NADP-ME expression construct (Detarsio et al., 2003) was used. Point mutations were introduced into the protein-coding region by performing site-directed mutagenesis with the following "back-to-back" primers: 5´-TCGTCCACCAGCCAAAC-3´ and 5´-AAAGGGTTTGATTGTTGACTCTCG-3´ ("for" and "rev" primers for the introduction of the S419E mutation; resulting construct has been termed pET32a*::*S419E NADP-ME) and 5´-GCGTCCACCAGCCAAACC-3´ as the "for" primer and the same "rev" primer as used in case of the first mutagenesis for the introduction of the S419A mutation (resulting construct has been termed pET32a*::*S419A-NADP-ME). All primers used for the mutagenesis procedure were phosphorylated at the 5´-terminus.

Site-directed mutagenesis has been performed as follows: usually, 5 similar 50 μL PCR reactions were performed simultaneously. Obtained DNA was pulled together and (after digestion from the agarose gel, if needed) subjected to the self-ligation by using the T4 DNA ligase (ThermoFisher Scientific). Self-ligation was performed by incubation of the mix at 22°C for 1 h. Subsequently, inactivation of the ligase by incubation of reaction mix at 65 °C for 10 min has been performed. In order to destroy the methylated WT (sample) DNA, the reaction mix after self-ligation has been subjected to the digestion at 37°C for 2 h by the DpnI (Thermo Scientific). After digestion, the DpnI has been inactivated by incubation at 80°C for 5 min. Concluding, obtained constructs were used for the transformation of the chemically-competent *E. coli* DH5α cells. After plasmid preparation, the success of the site-directed mutagenesis procedure has been confirmed by sequencing with three primers covering the whole length of the protein-coding region.

Expression and purification of recombinant ZmC$_4$-NADP-ME

The WT and two obtained mutated pET32*::*ME constructs were used for the transformation of the *E. coli* strain Rosetta (DE3) (Novagen). For the heterologous protein production, transformed cells were grown in 400 mL LB medium at 37 °C and agitation of 110 rpm in the

presence of 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} chloramphenicol until an OD$_{600}$ of 0.6 -0.8 was reached. To induce protein expression, final concentration of 1 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) was added to the culture and cells were grown for additional 20h under the same conditions. Harvest of the cells was performed by centrifugation at 4,000 g for 15 min. Obtained pellets were transferred into 50 mL Falcon tubes and were stored at -20°C until further usage. For the protein extraction, pellets were thawed on ice and resuspended in 20 mM Tris-HCl (pH 8.0) containing 500 mM NaCl, 5 mM imidazole, 2 mM phenylmethanesulfonyl fluoride (PMSF) and a spatula-tip amount of lysozyme, sonicated and centrifuged at 14,000 g for 20 min at 4°C to remove cell debris. Obtained supernatant has been deployed for protein purification using gravity-flow immobilized metal ion chromatography on nickel-nitrilotriacetic acid agarose (Ni-NTA Agarose, Qiagen). Prior to the supernatant loading, the column was equilibrated with 20 mM Tris-HCl buffer containing 500 mM NaCl and 5 mM imidazole. After the supernatant was loaded, columns were washed in four steps with 20 mM Tris-HCl (pH 8.0) and 500 mM NaCl buffers containing increasing concentrations of imidazole (5, 30, 40 and 50 mM) in order to isolate the His-tagged NADP-MEs. Elution was performed using four times 500 µL of 20 mM Tris-HCl buffer containing 500 mM NaCl and 300 mM imidazole. Eluted protein from the first elution fraction has been used for further kinetic measurements.

For all constructs used in this work, the calculated molecular mass of the expressed protein corresponded to the expected molecular mass of the fusion protein as follows: mature ZmC$_4$-NADP-ME (63.4 kDa) plus 17.3 kDa encoded by the expression vector.

Protein quantification, gel electrophoresis and immunological detection

Protein concentration was determined using the Pierce^{TM} BCA Protein Assay Kit (Thermo Fischer Scientific). SDS-PAGE was performed using 12% (w/v) polyacrylamide gels according to Laemmli (Laemmli, 1970). Proteins were visualized by staining with Coomassie Brilliant Blue or electroblotted onto a nitrocellulose membrane (Thermo Scientific) for immunological detection. The membranes were incubated for at least 1 h with the 1:7500 dilution of Anti-His-Tag antibody coupled to horseradish peroxidase (Anti-His-HRP; Miltenyi Biotec). After several washing step the membranes were incubated with a 1:2500 dilution of the Goat Anti-Rabbit IgG Antibody HRP-conjugate (Merck). The chemiluminescence signal was detected with Immobilon^{TM} Western Chemiluminescent HRP Substrate (Merck Millipore)

with subsequent visualization on an LAS-4000 Mini Luminescent Image Analyzer (GE Healthcare Life Sciences formerly Fuji).

Determination of kinetic parameters

NADP-ME activity was determined using a Synergy HT Biotek® Plate Reader system by measuring the formation of NADPH at 340 nm (extinction coefficient = 6.22 mM^{-1} cm^{-1}) at room temperature. The standard assay medium contained 0.5 mM NADP$^+$, 10 mM MgCl$_2$, 4 mM L-malate, 50 mM Tris-HCl (pH 8.0) and 0.1 - 0.8 µg enzyme per well in a final volume of 200 µL. The dependence of the activity with the pH of the medium was determined with the standard assay medium using different buffer systems: 50 mM MES (pH 5.5-6.5), 50 mM Tricine-Mops (pH 7.0-7.5), and 50 mM Tris-HCl (pH 7.5-9.0). The Michaelis (K_m in case of WT and S419A) and apparent Michaelis (K_{mapp} in case of S419E) constant of the substrate were determined by varying the concentration of one substrate while keeping the other components constant at fixed concentrations as described for the standard assay medium. NADP concentrations were varied between 2 and 150 µM in case of WT, between 10 and 800 µM in case of S419A and between 10 µM and 5 mM in case of S419E. Malate concentrations were varied between 0.001 and 8 mM in case of WT and S419A and between 0.1 and 10 mM in case of the S419E. As S419 was not saturated with NADP at least until 5 mM NADP, the K_m NADP for S419E could not be estimated. Measuremets of enzymatic activity using high NADP concentrations are technicaly impair by the high absortion of this compound; thus, other kinetik parameters of S419E, apparent parameters (k_{catapp} and K_{mapp} malate) were estimated using 0.5 mM NADP. All kinetic parameters were calculated using at least three biological replicates and adjusted to non-lineal regression. Data were fitted with the Prism 6 (GraphPad Software).

Circular Dichroism

All recombinant versions of ZmC$_4$-NADP-ME were subjected to circular dichroism (CD) analysis in 20 mM NaP$_i$, pH 8.0 or 7.0; 5 mM MgCl$_2$. Protein concentration was determined at 280 nm using a Jasco V-650 spectrophotometer. The protein concentration (c, g L^{-1}) was calculated by the equation $c = A_{280} \cdot Mw/(\varepsilon \cdot d)$, in which ε represents the molar extinction coefficient at 280 nm (76250 L mol^{-1} cm^{-1}), d is the cell path (1 cm) and Mw is molecular weight (80.7 kDa for all three proteins).

CD spectra between 240 and 190 nm were obtained averaging 10 repetitive scans in a Jasco J-810 spectropolarimeter. Mean residue ellipticity ($[\Theta]$, deg cm^2 dmol^{-1}) was obtained by the equation $[\Theta] = MRW \cdot \Theta/(10 \cdot c \cdot d)$, where MRW (the mean amino acid residue weight) was calculated as relation of the protein's molecular weight (Mw) to the number of peptide bonds ($N - 1$; where N is the number of amino acids in the chain), d is the cell path (0.1 cm), c is the protein concentration and Θ is the observed ellipticity in millidegrees (machine units). Determination of the protein secondary structure contents was performed using the CONTIN-LL method (van Stokkum et al., 1990) and the reference dataset 4 (Sreerama and Woody, 2000) at the Dichroweb online server (http://dichroweb.cryst.bbk.ac.uk) (Whitmore and Wallace). Prism Software (http://www.graphpad.com/scientific-software/prism/) was used for visualization of the deconvolution by Dichroweb.

Analytical Ultracentrifugation

Sedimentation velocity experiments were carried out using a Beckman Optima XL-A Analytical Ultracentrifuge. All recombinant versions of ZmC$_4$-NADP-ME were assayed in 10 mM Tris-HCl buffer (pH 7.0 and 8.0) with addition of 5 mM MgCl$_2$. Buffer exchange procedure was performed using the Amicon Ultra 0.5 mL centrifugal units with nominal molecular weight limit of 50 kDa (Merck) and final protein concentration was adjusted to 0.6 mg mL^{-1}. Samples (230 µg) and corresponding buffer solutions (400µL) were loaded into aluminium double sector centerpieces separately and built up in a Beckman An-50 Ti rotor. Experiments were performed at 20 °C and a rotor speed of 35,000 rpm. Protein samples were monitored by UV absorbance at 280 nm in a continuous mode with a radial resolution of 0.003 cm. In time intervals of about 2 min scans of the radial concentration profile were collected until the protein was fully sedimented. Data were analysed using the c(s) model in the software package SEDFIT (Schuck and Rossmanith, 2000). For data analysis, a resolution of 0.1 S with a confidence level (F-ratio) of 0.95 was chosen for the appropriate s-value range within 0 to 30.0 S. Density and viscosity of the solvent had been calculated with the software Sednterp from tabulated values; $\rho = 0.99885$ g cm^{-3} and $\eta = 0.01006$ g cm^{-1}s^{-1}. The protein partial specific volume $\bar{v}$ applied for WT and mutants was 0.7387 cm³ g^{-1}. Sedimentation coefficients are reported as $s_{20, w}$ values, i.e. normalized to 20°C and water as a solvent. Graphic output was generated by Gussi (Version 1.2.1) (Brautigam, 2015) and the final sedimentation coefficient distribution was normalized based on the maximum peak height.

NADP-ME activity measurements in extracts of maize leaves

Soluble protein extracts were prepared by homogenizing 0.25 g of frozen maize leaves in 500 µL of ice-cold extraction buffer containing 100 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 0.1 mM EDTA, 10% (v/v) glycerol, 2 mM DTT, 1 mM PMSF. The homogenate was clarified by centrifugation at 20,000 g during 20 min at 4°C. Protein concentration of the protein extract was determined using the Amido Black assay (Schaffner and Weissmann, 1973). NADP-ME activity was measured in medium containing 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 0.5 mM NADP, 10 mM L-malate and 1.5 - 2.2 µg protein extract in a final volume of 0.2 mL. The reaction was started by addition of $NADP^+$. Measurements were conducted using three independent replicates.

Gas Chromatography - Mass Spectrometry (GC-MS) analysis of primary metabolites

Frozen maize leaves were ground in liquid nitrogen until a fine powder was obtained. Ground material (around 50 mg) was extracted in methanol:chloroform:water (5:2:2) with ribitol for internal standardization as described by Lee and Fiehn (Lee and Fiehn, 2008).

An aliquot of the supernatant (150 µl) was dried under vacuum, and the residue was derivatized for 120 min at 37°C (in 60 µl of 30 mg ml^{-1} methoxyamine hydrochloride in pyridine) followed by a 30 min treatment at 37°C with 120 µl of MSTFA. The GC-MS system used was a gas chromatograph coupled to a time-of-flight mass spectrometer (Leco Pegasus HT TOF-MS). An auto sampler Gerstel Multi-Purpose system injected the samples. Helium was used as carrier gas at a constant flow rate of 2 mL s^{-1} and gas chromatography was performed on a 30 m DB-35 column. The injection temperature was 230°C and the transfer line and ion source were set to 250°C. The initial temperature of the oven (85°C) increased at a rate of 15°C/min up to a final temperature of 360°C. After a solvent delay of 180 s mass spectra were recorded at 20 scans s-1 with m/z 70-600 scanning range. Chromatograms and mass spectra were evaluated by using Chroma TOF 4.5 (Leco) and TagFinder 4.2 software (Ishizaki et al., 2005;Lisec et al., 2006).

Structural analysis of SbC$_4$-NADP-ME

Recombinant SbC4-NADP-ME was expressed in *E. coli* and isolated from the bacterial extract as described in (Saigo et al., 2013). The His-tag of the protein was removed by incubation with 1:20 (thrombin:SbC4-NADP-ME) for 2 h at 16°C. The cleaved His-tag was eliminated by size exclusion chromatography. The eluted protein fractions were combined and subjected to concentration and buffer exchange using 10 mM Tris-HCl (pH 8.0) containing 5 mM MgCl2 employing a Vivaspin® 20 centrifugal concentrator (Sartorius AG). For crystallization studies the protein concentration of the sample was adjusted to 6 mg mL^{-1}.

Initial identification of crystallization conditions was carried out using the sitting-drop vapor diffusion method, with a robotic workstation (Honeybee963®, Isogen Life Science). Sitting drops were set using 400 nL of a 1:1 mixture of protein and mother liquor, equilibrating against 150 μL mother liquors in the reservoir of Greiner plates, both at 4°C and 20°C. The mother liquors screened were from commercial kits: Crystal screens HR2-110 and HR-112 (Hampton Research) and PEG Suite solutions (QIAGEN). The best crystal hit contained 0.2 M KF and 20% PEG 3350 as mother liquor. The crystal growth habit comparing both crystallization temperatures was similar, so we continued working only at 20°C. Initial SbC4-NADP-ME crystals were manually optimized by varying precipitant and protein concentrations using VDX plates (Hampton Research) with a hanging-drop setup. The best crystallization condition was ultimately achieved using 0.2 M KF and 30% PEG 3350 reservoir solution at 20°C, with protein at 5.0 mg/mL in 20 mM Tris pH 8.0, 50 mM NaCl, 2 mM NADP, 40 mM pyruvate and 10 mM MgCl2. Crystals appeared in about 7 days. Cryo-protection was achieved by slowly adding cryo-protection solution (25% (v/v) glycerol, 30% (w/v) PEG 3350, 0.1 M Tris-HCl pH 8.0, 0.2 M MgCl2, 2 mM NADP and 40 mM pyruvate) to the drop in small volumes (~ 5% drop volume step-wise, until reaching >15% glycerol), then rapidly soaked in 100% cryo-protection solution, and flash-cooled in liquid N2 for storage until data collection. X-ray diffraction data collection was performed with a Micromax007-HF generator (Rigaku) with a rotating Cu anode, Varimax-HF (Rigaku) optics and a Mar345 image plate detector (Mar Research) reaching 2.0 Å resolution.

All data obtained for the SbC4-NADP-ME was processed with XDS (Kabsch, 2010) or Mosflm (Battye et al., 2011) and scaled with aimless (Evans and Murshudov, 2013). Phasing was performed by MR using chain A from the human mitochondrial malic enzyme (pdb-code: 1DO8) as template. The resulting structure was used as starting point for further model building. The structure was then refined by iterative cycles of manual refinement using Coot (Emsley et al., 2010) and Refmac5 (Murshudov et al., 2011) from CCP4 suite (Collaborative

Computational Project, 1994) or Buster (Bricogne et al., 2009). Although chain C and D are not unambiguously traceable we used all (21838) atoms for the refinement. The structures were deposited in the protein data bank (wwpdb) under the accession code 6C7N.

Generation of the complex structures for molecular dynamics simulations

For the preparation of the complex structures, MOE (MOE, 2018) was used to superimpose an in-house low resolution crystal structure of ZmC_4-NADP-ME with the crystal structures of ME from pigeon liver (PDB accession code 1GQ2) and SbC_4-NADP-ME (PDB accession code 6C7N) according to the C_α-RMSD of residues adjacent to the NADP binding site. This overlay was used for the initial placement of NADP and a Mn^{2+} ion in the active site of ZmC_4-NADP-ME by copying the respective atom coordinates of Mn^{2+} from the pigeon liver ME and of NADP from SbC_4-NADP-ME, respectively. The resulting complex was prepared using MOE's Protein Preparation Tool. Finally, the sidechain of residue S419 of wild type ZmC_4-NADP-ME was mutated to glutamate, to obtain a model for the variant.

Molecular dynamics simulations of NADP/ZmC_4-NADP-ME complexes

System setup

The following procedures were performed using the Amber18 software package (Case et al., 2018). We used tLEaP to solvate each of the complexes in a periodic box of TIP3P water (Jorgensen et al., 1983), in such a way that the distance between the edges of the box and any solute atoms is at least 11 Å. The system was neutralized by replacing five random water molecules by sodium ions. The ff14SB force field (Hornak et al., 2006) was used for parameterizing the protein, and Mn^{2+} parameters were taken from the 12-6 LJ set of ion parameters by Li and Merz (Li et al., 2013). Partial charges according to the RESP procedure (Bayly et al., 1993) for NADP were derived from the R.E.DD.B. (Dupradeau et al., 2008) project F-91, while for bond parameters the GAFF2 force field (Case et al., 2018) was used.

Thermalization protocol

All simulations were carried out using the GPU version of pmemd (Le Grand et al., 2013;Case et al., 2018) and a time step of 2 fs. The Langevin thermostat (Pastor et al., 1988) was used for temperature control with a collision frequency of $\gamma = 2.0$ ps^{-1}. For treatment of long-range electrostatic interactions the Particle Mesh Ewald method (Darden et al., 1993) was used with a cutoff of 8.0 Å. The SHAKE algorithm (Ryckaert et al., 1977) was used to constrain bond lengths involving hydrogen atoms.

Initially, an energy minimization was performed on solvent and solute atoms independently for 3,000 steps each with the steepest decent algorithm, followed by 2,000 steps with the conjugate gradient algorithm. Thereafter, all atoms were energy-minimized together for 7,000 steps of the steepest decent algorithm and 3,000 steps of the conjugate gradient algorithm. Then, the systems were heated up to 300 K at a constant heating rate over 20 ps of NVT-MD and simulated for further 5 ps at 300 K (with positional restraints on the atoms of the NADP-enzyme complex with a force constant of 2 kcal·mol^{-1}·Å^{-2}). For density adaptation, 75 ps of NPT-MD were performed with the atoms of the NADP-enzyme complex restrained with a force constant of 2 kcal·mol^{-1}·Å^{-2} and a pressure relaxation time of $\tau_p = 1.0$ ps, followed by 1,700 ps of NPT-MD with the same restraints and $\tau_p = 2.0$ ps. Another ten iterations of minimization were performed on the system with iteratively decreasing restraints on the solute atoms (force constants ranging from 2 to 0 kcal·mol^{-1}·Å^{-2}), with 500 steps of steepest decent minimization and 500 steps of conjugate gradient each. Finally, the systems were heated up again to 300 K over 100 ps of NVT-MD, followed by 3,100 ps of unrestrained NVT-MD.

Production protocol

For the WT and the variant system, 300 ns of unbiased NVT-MD were performed each. Coordinates for post-processing and further analysis were extracted every 100 ps. For trajectory post-processing, we used CPPTRAJ (Roe and Cheatham, 2013).

Calculation of effective binding free energies

The following calculations were performed using mm_pbsa.pl (Gohlke et al., 2003) from AmberTools18 (Case et al., 2018). For both systems, the change of the effective energy due to the binding of NADP was calculated by the single-trajectory MM-GB/SA approach (Kuhn and Kollman, 2000;Hou et al., 2011). All counter ions and water molecules were stripped from the trajectory, and the sum of internal, electrostatic, and van der Waals energies were calculated for each snapshot using the above-mentioned force fields. The polar fraction of the solvation free energy was calculated using the OBC^{II} generalized Born model (Onufriev et al., 2004) with mbondi2 radii. The dielectric constants were set to 1 for the solute and 80 for the solvent. Non-polar contributions were calculated by a solvent-accessible surface area-dependent term applying a surface tension of $\gamma = 0.0072$ kcal·mol^{-1}·Å^{-2}. Contributions due to changes in the configurational entropy of the ligand or the receptor upon complex formation were neglected, in order to avoid introducing uncertainty in the computations (Gohlke and Case, 2004;Weis et al., 2006;Hou et al., 2011).

Accession Numbers

The structure of SbC$_4$-NADP-ME was deposited in the protein data bank (wwpdb) under the accession code 6C7N.

Supplemental data

Supplemental Figure 1. Alignment of plastidic C$_4$- and nonC$_4$-NADP-ME sequences in Poaceae showing conservation of Ser419 and the BARK substrate motif (DSKGL). We obtained the amino acid sequences of the plastidic C$_4$- and nonC$_4$-NADP-ME isoforms of *Z. mays* cultivar Ensembl-18 (Sequence IDs: GRMZM2G085019_T01, GRMZM2G122479_P01) and *S. bicolor* cultivar v3.1.1 (Sequence IDs: Sobic.003G036200.1, Sobic.009G108700.1) from the Phytozome 11 database. We used the maize C$_4$ and nonC$_4$-NADP-ME sequences as BLAST queries to identify the corresponding orthologs in the *Setaria italica* cultivar v2.2 proteome in the same database (Sequence ID: Seita.5G134300.1). We also used the maize nonC$_4$-NADP-ME sequence as a BLAST query to identify the unique plastidial NADP-ME in the *Oryza sativa* cultivar v7_JGI proteome (LOC_Os01g09320.1). All sequences were aligned using PRANK. The asterisks (*) denote the position of Ser419. The green box indicates the BARK substrate motif, which includes the Ser419, among plastidic photosynthetic and non-photosynthetic NADP-ME isoforms within the Poaceae. Zm: *Zea mays*; Sb: *Sorghum bicolor*; Si: *Setaria italica*; Os: *Oryza sativa*.

Supplemental Figure 2. SDS-PAGE stained with Coomassie-Blue (lines 1-4) and analysed by immunoblot (line 5) of protein fractions during the isolation of recombinant ZmC$_4$-NADP-ME variants. M, molecular weight markers (Spectra Multicolor Broad range Protein Ladder (Thermo Fisher Scientific); 1, 20 µg of non-induced cell culture lysate; 2, 20 µg of cell culture lysate 16 h after induction of protein production; 3, 20 µg of crude extract cleared by centrifugation; 4 and 5, 1.5 µg of affinity purified protein; The size of the recombinant proteins (80.7 kDa; mature ZmC$_4$-NADP-ME 63.4 kDa plus 17.3 kDa encoded by the expression vector) is indicated on the left. Anti-His-HRP conjugate antibodies were used for the immunoblots.

Supplemental Figure 3. Cartoon and surface representation of SbC$_4$-NADP-ME chain B. The bound cofactor NADP and the phosphorylation site Ser419, located on the surface of the

enzyme, are highlighted as sticks.

Supplemental Table 1. Secondary structure contents of ZmC4-NADP-ME WT, S419A and S419E estimated from the CD-spectra obtained at pH of 7.0 and 8.0. The Dichroweb server (http://dichroweb.cryst.bbk.ac.uk) (Whitmore and Wallace, 2004) was used to calculate protein secondary structure contents from experimentally acquired data. The secondary structures are presented as a sum of regular α-helix, distorted α-helix, regular β-strand, distorted β-strand, turns and disordered conformations in case of each analysed protein. The normalized root-mean-square deviations, a measure for the quality of the data evaluation expressed as the difference between the original CD spectrum and its reconstruction, were always lower than 0.1.

Supplemental Table 2. Changes in metabolites in maize leaves. Material was harvested at ZT0.5, 2, 4, 8, 15.5, 16.5, and 23.5 (the last two time points represent the beginning and end of night period, respectively). Values presented are mean ± standard error of three replicates. Bold letters indicate significant differences by Student´s t test (P< 0.05) compared to the ZT23.5 time point.

Supplemental Table 3. Protein structure data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

Supplemental Table 4. Hydrogen bonds with and distances to the cofactor NADP in monomer A of SbC4-NADP-ME.

and IML is a recipient of FCT doctorate fellowship (PD/BD/113982/2015 within MolBioS PD/00133/2012). LC-MS/MS analyses were performed at the Mass Spectrometry Unit (UniMS) facility of iBET/ITQB-UNL, Oeiras, Portugal. We are grateful for computational support by the "Zentrum für Informations und Medientechnologie" at the Heinrich Heine University and the computing time provided by the John von Neumann Institute for Computing (NIC) to H.G. on the supercomputer JURECA at Jülich Supercomputing Centre (JSC) (user ID: HKF7).

AUTHOR CONTRIBUTIONS

Conceptualization, V.G.M.; Methodology, V.G.M. and I.A.A.; Investigation, A.B., B.M.A., C.E.A., I.M.L., A.H., C.D., F.T., A.B., A.A., D.B. and H.G.; Writing – Original Draft, V.G.M., I.A.A., A.B., A.H., C.E.A., M.F.D., L.N-S. and H.G.; Funding Acquisition, V.G.M. and I.A.A.; Resources, V.G.M., I.A.A., M.F.D, A.F; Supervision, V.G.M., I.A.A., M.F.D.

FIGURE LEGENDS

Figure 1. TripleTOF 6600 tandem MS data of the phosphopeptide acVWLVDpSK of ZmC$_4$-NADP-ME. The detected b (N-terminal, in red) and y (C-terminal, in blue) fragment ions are labeled in the spectrum. Ac denotes N-terminus acetylation and pS denotes phosphorylated serine. Precursor charge: +2; monoisotopic m/z: 484.7265 Da (-1.60 mmu/-3.30 ppm). Confidence (ProteinPilot): 96.4% (Confidence threshold for FDR≤1% was equal to 93.7%).

Figure 2. Biochemical characterization of recombinant ZmC$_4$-NADP-ME versions. A, Dependence of the activity on the pH of the assay. The values represent the mean ± standard error of at least four independent enzyme preparations, each measured in triplicate. **B**, Kinetic parameters at pH 8.0. Kinetic data were best fitted by nonlinear regression analysis. The values represent the mean ± standard error of at least three independent enzyme preparations, each measured in triplicate.

Figure 3. Exploration of the secondary structure organization and quaternary composition of recombinant ZmC$_4$-NADP-ME versions. A and B: CD spectra of ZmC$_4$-NADP-ME WT and the two mutated versions obtained at pH 7.0 (A) and 8.0 (B). Ten

accumulations each were collected from 190 to 260 nm for 0.16 mg mL^{-1} enzyme, at 20°C. Each graph is showing the reconstructed curves obtained by applying CONTIN/LL algorithm for data evaluation as provided by the Dichroweb server. **C and D**: Continuous sedimentation coefficient distribution of ZmC$_4$-NADP-ME WT and the mutated versions at pH 7.0 (C) and 8.0 (D). Data was fitted with the ls-g*(s) model in the software package SEDFIT.

Figure 4. Measurements of different parameters in extracts of maize leaves during a diurnal cycle. A, Quantification of total ZmC$_4$-NADP-ME protein by SWATH-MS. **B**, Quantification of the phosphopeptide VWLVDpSK, corresponding to the phosphorylation of ZmC$_4$-NADP-ME at S419. **C**, Total NADP-ME activity measured with 1.5 - 2.0 µg protein extracts. **D**, Malate concentration assayed by GC-MS analysis. Values represent mean ± standard error of four (A and B) or three (C and D) biological replicates. The dark period is highlighted in grey. Statistical analyses were performed against the first time point in the night (ZT16.5). Letters indicate that the value is statistically significant at a, 0.05; b, 0.01; c, 0.001 levels.

Figure 5. Heat map showing the changes in metabolite relative levels. Metabolites were sorted based on the sum across the time points and then hierarchical cluster was applied. The score, depicted in blue, represents decreased levels, while red color indicates increased levels. The depth of the color corresponds to the magnitude of the change in levels. ZT, Zeitgeber Time. The log2 fold change scale is indicated below the heat map. For details, see Table S2.

Figure 6. SbC$_4$-NADP-ME crystal structure and amino acids involved in cofactor binding. A, Cartoon and ribbon representation of SbC$_4$-NADP-ME. Monomer A (red) and B (blue) with bound cofactor NADP (colored spheres). Large parts of chain C and D are not well resolved by electron density; therefore, we depicted these chains as gray ribbons only for clarity. **B.** Monomer A from SbC$_4$-NADP-ME with the bound cofactor NADP (represented by sticks). **C.** Residues involved in cofactor binding are depicted as sticks with labels. Hydrogen bonding pattern in the SbC$_4$-NADP-ME cofactor binding site (distances are omitted for clarity and are listed in Table S4).

Figure 7. Root mean square deviations and effective binding energies of NADP/ZmC$_4$-NADP-ME complexes. A. C$_\alpha$-RMSD of WT ZmC$_4$-NADP-ME (blue) and the S419E variant (orange) along the MD trajectories. **B.** RMSD of NADP in the binding site after fitting on the enzyme; color code as in panel A. **C.** Time series of the effective binding energies for NADP

binding to WT ZmC4-NADP-ME (blue) and the S419E variant (orange); on the right, the probability distributions of the effective binding energies are shown. **D.** Various binding poses of NADP in the S419E variant at indicated time points in the trajectory; the mutated residue E419 is highlighted in yellow.
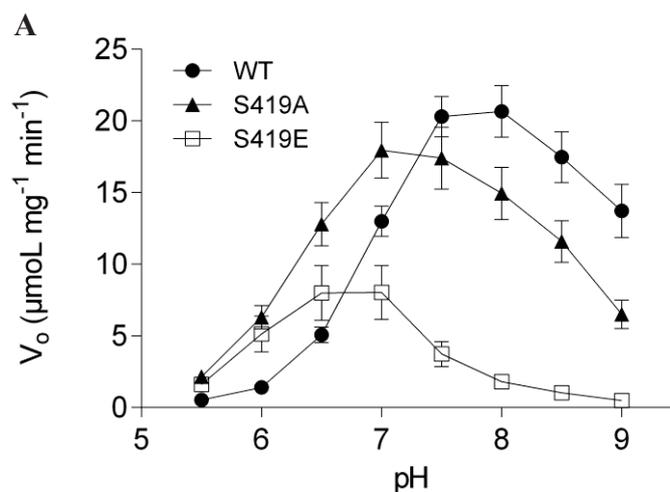
REFERENCES

Alvarez, C.E., Detarsio, E., Moreno, S., Andreo, C.S., and Drincovich, M.F. (2012). Functional characterization of residues involved in redox modulation of maize photosynthetic NADP-malic enzyme activity. *Plant Cell Physiol* 53**,** 1144-1153.

Alvarez, C.E., Saigo, M., Margarit, E., Andreo, C.S., and Drincovich, M.F. (2013). Kinetics and functional diversity among the five members of the NADP-malic enzyme family from Zea mays, a C4 species. *Photosynth Res* 115**,** 65-80.

Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S.G., and Pandey, A. (2007). A curated compendium of phosphorylation motifs. *Nat Biotechnol* 25**,** 285-286.

Arrivault, S., Obata, T., Szecowka, M., Mengin, V., Guenther, M., Hoehne, M., Fernie, A.R., and Stitt, M. (2017). Metabolite pools and carbon flow during C4 photosynthesis in maize: 13CO2 labeling kinetics and cell type fractionation. *J Exp Bot* 68**,** 283-298.

Bailey, K.J., Gray, J.E., Walker, R.P., and Leegood, R.C. (2007). Coordinate regulation of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxykinase by light and CO2 during C4 photosynthesis. *Plant Physiol* 144**,** 479-486.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67**,** 271-281.

Bayly, C.I., Cieplak, P., Cornell, W.D., and Kollman, P.A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry* 97**,** 10269-10280.

Bellasio, C., and Griffiths, H. (2014). Acclimation to low light by C4 maize: implications for bundle sheath leakiness. *Plant Cell Environ* 37**,** 1046-1058.

Brautigam, C.A. (2015). Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol* 562**,** 109-133.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, P., Roversi, P., Sharff, A., Smart, O., Vonrhein, C., and Womack, T. (2009). BUSTER version 2.9. . *Cambridge, United Kingdom: Global Phasing Ltd.*

Burnell, J.N., and Hatch, M.D. (1985). Regulation of C4 photosynthesis: purification and properties of the protein catalyzing ADP-mediated inactivation and Pi-mediated activation of pyruvate,Pi dikinase. *Arch Biochem Biophys* 237**,** 490-503.

Case, D.A., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., III, T.E.C., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Ghoreishi, D., Gilson, M.K., Gohlke, H., Goetz, A.W., Greene, D., Harris, R., Homeyer, N., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D.J., Merz, K.M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Smith, J., Salomon-Ferrer, R., Swails, J., Walker, R.C., Wang, J., Wei, H., Wolf, R.M., Wu, X., Xiao, L., York, D.M., and Kollman, P.A. (2018). AMBER 2018. *University of California, San Francisco.*

Chao, Q., Liu, X.Y., Mei, Y.C., Gao, Z.F., Chen, Y.B., Qian, C.R., Hao, Y.B., and Wang, B.C. (2014). Light-regulated phosphorylation of maize phosphoenolpyruvate carboxykinase plays a vital role in its activity. *Plant Mol Biol* 85**,** 95-105.

Chastain, C.J., Botschner, M., Harrington, G.E., Thompson, B.J., Mills, S.E., Sarath, G., and Chollet, R. (2000). Further analysis of maize C-4 pyruvate,orthophosphate dikinase phosphorylation by its bifunctional regulatory protein using selective substitutions of the regulatory Thr-456 and catalytic His-458 residues. *Archives of Biochemistry and Biophysics* 375**,** 165-170.

Chen, Y.B., Lu, T.C., Wang, H.X., Shen, J., Bu, T.T., Chao, Q., Gao, Z.F., Zhu, X.G., Wang, Y.F., and Wang, B.C. (2014). Posttranslational Modification of Maize Chloroplast Pyruvate Orthophosphate Dikinase Reveals the Precise Regulatory Mechanism of Its Enzymatic Activity. *Plant Physiol* 165**,** 534-549.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50**,** 760-763.

Collins, B.C., Gillet, L.C., Rosenberger, G., Rost, H.L., Vichalkovski, A., Gstaiger, M., and Aebersold, R. (2013). Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat Methods* 10**,** 1246-1253.

Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* 98**,** 10089-10092.

Detarsio, E., Maurino, V.G., Alvarez, C.E., Muller, G.L., Andreo, C.S., and Drincovich, M.F. (2008). Maize cytosolic NADP-malic enzyme (ZmCytNADP-ME): a phylogenetically distant isoform specifically expressed in embryo and emerging roots. *Plant Mol Biol* 68**,** 355-367.

Detarsio, E., Wheeler, M.C., Campos Bermudez, V.A., Andreo, C.S., and Drincovich, M.F. (2003). Maize C4 NADP-malic enzyme. Expression in Escherichia coli and characterization of site-directed mutants at the putative nucleoside-binding sites. *J Biol Chem* 278**,** 13757-13764.

Dissmeyer, N., and Schnittger, A. (2011). Use of phospho-site substitutions to analyze the biological relevance of phosphorylation events in regulatory networks. *Methods Mol Biol* 779**,** 93-138.

Dupradeau, F.Y., Cezard, C., Lelong, R., Stanislawiak, E., Pecher, J., Delepine, J.C., and Cieplak, P. (2008). REDDB: A database for RESP and ESP atomic charges, and force field libraries. *Nucleic Acids Research* 36**,** D360-D367.

Emms, D.M., Covshoff, S., Hibberd, J.M., and Kelly, S. (2016). Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C4 Species. *Mol Biol Evol* 33**,** 1796-1806.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66**,** 486-501.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69**,** 1204-1214.

Fahnenstich, H., Saigo, M., Niessen, M., Zanor, M.I., Andreo, C.S., Fernie, A.R., Drincovich, M.F., Flugge, U.I., and Maurino, V.G. (2007). Alteration of organic acid metabolism in Arabidopsis overexpressing the maize C(4)NADP-malic enzyme causes accelerated senescence during extended darkness. *Plant Physiology* 145**,** 640-652.

Furbank, R.T. (2011). Evolution of the C4 photosynthetic mechanism: are there really three C4 acid decarboxylation types? *Journal of Experimental Botany* 62**,** 3103-3108.

Furbank, R.T., and Hatch, M.D. (1987). Mechanism of c(4) photosynthesis: the size and composition of the inorganic carbon pool in bundle sheath cells. *Plant Physiol* 85**,** 958-964.

Furbank, R.T., Jenkins, C.L.D., and Hatch, M.D. (1990). C4 photosynthesis—quantum requirement, C4 acid overcycling and Q-cycle involvement. *Australian Journal of Plant Physiology* 17**,** 1-7.

Gohlke, H., and Case, D.A. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *Journal of Computational Chemistry* 25**,** 238-250.

Gohlke, H., Kiel, C., and Case, D.A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *Journal of Molecular Biology* 330**,** 891-913.

Hatch, M.D. (1987). C-4 Photosynthesis - a Unique Blend of Modified Biochemistry, Anatomy and Ultrastructure. *Biochimica Et Biophysica Acta* 895**,** 81-106.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* 65**,** 712-725.

Hou, T.J., Wang, J.M., Li, Y.Y., and Wang, W. (2011). Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* 51**,** 69-82.

Iglesias, A.A., and Andreo, C.S. (1990). Kinetic and Structural Properties of NADP-Malic Enzyme from Sugarcane Leaves. *Plant Physiol* 92**,** 66-72.

Ishizaki, K., Larson, T.R., Schauer, N., Fernie, A.R., Graham, I.A., and Leaver, C.J. (2005). The critical role of Arabidopsis electron-transfer flavoprotein:ubiquinone oxidoreductase during dark-induced starvation. *Plant Cell* 17**,** 2587-2600.

Jiao, J.A., and Chollet, R. (1988). Light/dark regulation of maize leaf phosphoenolpyruvate carboxylase by in vivo phosphorylation. *Arch Biochem Biophys* 261**,** 409-417.

Jiao, J.A., and Chollet, R. (1989). Regulatory seryl-phosphorylation of C4 phosphoenolpyruvate carboxylase by a soluble protein kinase from maize leaves. *Arch Biochem Biophys* 269**,** 526-535.

Jiao, J.A., and Chollet, R. (1990). Regulatory phosphorylation of serine-15 in maize phosphoenolpyruvate carboxylase by a C4-leaf protein-serine kinase. *Arch Biochem Biophys* 283**,** 300-305.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* 79**,** 926-935.

Kabsch, W. (2010). Xds. *Acta Crystallogr D Biol Crystallogr* 66**,** 125-132.

Kromdijk, J., Ubierna, N., Cousins, A.B., and Griffiths, H. (2014). Bundle-sheath leakiness in C4 photosynthesis: a careful balancing act between CO2 concentration and assimilation. *J Exp Bot* 65**,** 3443-3457.

Kuhn, B., and Kollman, P.A. (2000). Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *Journal of Medicinal Chemistry* 43**,** 3786-3791.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227**,** 680-685.

Lambert, J.P., Ivosev, G., Couzens, A.L., Larsen, B., Taipale, M., Lin, Z.Y., Zhong, Q., Lindquist, S., Vidal, M., Aebersold, R., Pawson, T., Bonner, R., Tate, S., and Gingras, A.C. (2013). Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat Methods* 10**,** 1239-1245.

Le Grand, S., Gotz, A.W., and Walker, R.C. (2013). SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications* 184**,** 374-380.

Lee, Y., and Fiehn, O. (2008). High quality metabolomic data for Chlamydomonas reinhardtii. *Plant Methods* 4**,** 7.

Li, P.F., Roberts, B.P., Chakravorty, D.K., and Merz, K.M. (2013). Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *Journal of Chemical Theory and Computation* 9**,** 2733-2748.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 1**,** 387-396.

Luis, I.M., Alexandre, B.M., Oliveira, M.M., and Abreu, I.A. (2016). Selection of an Appropriate Protein Extraction Method to Study the Phosphoproteome of Maize Photosynthetic Tissue. *PLoS One* 11**,** e0164387.

Molecular Operating Environment (MOE), 2016.8. *Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7*, 2018

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67**,** 355-367.

Onufriev, A., Bashford, D., and Case, D.A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins-Structure Function and Bioinformatics* 55**,** 383-394.

Pastor, R.W., Brooks, B.R., and Szabo, A. (1988). An Analysis of the Accuracy of Langevin and Molecular-Dynamics Algorithms. *Molecular Physics* 65**,** 1409-1419.

Pick, T.R., Bräutigam, A., Schlüter, U., Denton, A.K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U., and Weber, A.P.M. (2011). Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C4 Model and Provides Candidates for Regulation. *Plant Cell* 23**,** 4208-4220.

Roe, D.R., and Cheatham, T.E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* 9**,** 3084-3095.

Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics* 23**,** 327-341.

Sage, R.F., Sage, T.L., and Kocacinar, F. (2012). Photorespiration and the Evolution of C-4 Photosynthesis. *Annu Rev Plant Biol* 63**,** 19-47.

Saigo, M., Alvarez, C.E., Andreo, C.S., and Drincovich, M.F. (2013). Plastidial NADP-malic enzymes from grasses: unraveling the way to the C4 specific isoforms. *Plant Physiol Biochem* 63**,** 39-48.

Saigo, M., Bologna, F.P., Maurino, V.G., Detarsio, E., Andreo, C.S., and Drincovich, M.F. (2004). Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Mol Biol* 55**,** 97-107.

Schaffner, W., and Weissmann, C. (1973). A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal Biochem* 56**,** 502-514.

Schuck, P., and Rossmanith, P. (2000). Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* 54**,** 328-341.

Sharwood, R.E., Sonawane, B.V., and Ghannoum, O. (2014). Photosynthetic flexibility in maize exposed to salinity and shade. *J Exp Bot* 65**,** 3715-3724.

Sreerama, N., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287, 252-260.

van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R., and Groen, F.C. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem* 191, 110-118.

Walker, R.P., Chen, Z.H., Acheson, R.M., and Leegood, R.C. (2002). Effects of phosphorylation on phosphoenolpyruvate carboxykinase from the C4 plant Guinea grass. *Plant Physiol* 128, 165-172.

Walker, R.P., and Leegood, R.C. (1996). Phosphorylation of phosphoenolpyruvate carboxykinase in plants. Studies in plants with C4 photosynthesis and Crassulacean acid metabolism and in germinating seeds. *Biochem J* 317 (Pt 3), 653-658.

Weis, A., Katebzadeh, K., Soderhjelm, P., Nilsson, I., and Ryde, U. (2006). Ligand affinities predicted with the MM/PBSA method: Dependence on the simulation method and the force field. *Journal of Medicinal Chemistry* 49, 6596-6606.

Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* 32, W668-673.

Zell, M.B., Fahnenstich, H., Maier, A., Saigo, M., Voznesenskaya, E.V., Edwards, G.E., Andreo, C., Schleifenbaum, F., Zell, C., Drincovich, M.F., and Maurino, V.G. (2010). Analysis of Arabidopsis with highly reduced levels of malate and fumarate sheds light on the role of these organic acids as storage carbon molecules. *Plant Physiol* 152, 1251-1262.

Figure 1. TripleTOF 6600 tandem MS data of the phosphopeptide acVWLVDpSK of ZmC4-NADP-ME. The detected b (N-terminal, in red) and y (C-terminal, in blue) fragment

ions are labeled in the spectrum. Ac denotes N-terminus acetylation and pS denotes phosphorylated serine. Precursor charge: +2; monoisotopic m/z: 484.7265 Da (-1.60 mmu/-3.30 ppm). Confidence (ProteinPilot): 96.4% (Confidence threshold for FDR≤1% was equal to 93.7%).

A

B

ZmC$_4$-NADP-ME	WT	S419E	S419A
k_{cat} (s^{-1})	12.2 ± 1.4	2.1 ± 0.1*	8.3 ± 1.6
K_m NADP (μM)	12.3 ± 0.7	-	159.5 ± 7.5
K_m malate (mM)	0.19 ± 0.01	2.68 ± 0.14*	0.47 ± 0.08

-, non-determined, as saturation was not achieved. *Apparent kinetic parameters (k_{catapp} and K_{mapp} malate) of S419E were estimated using a non-saturating NADP

Figure 2. Biochemical characterization of recombinant ZmC$_4$-NADP-ME versions. A, Dependence of the activity on the pH of the assay. The values represent the mean ± standard error of at least four independent enzyme preparations, each measured in triplicate. **B**, Kinetic parameters at pH 8.0. Kinetic data were best fitted by nonlinear regression analysis. The values represent the mean ± standard error of at least three independent enzyme preparations, each measured in triplicate.
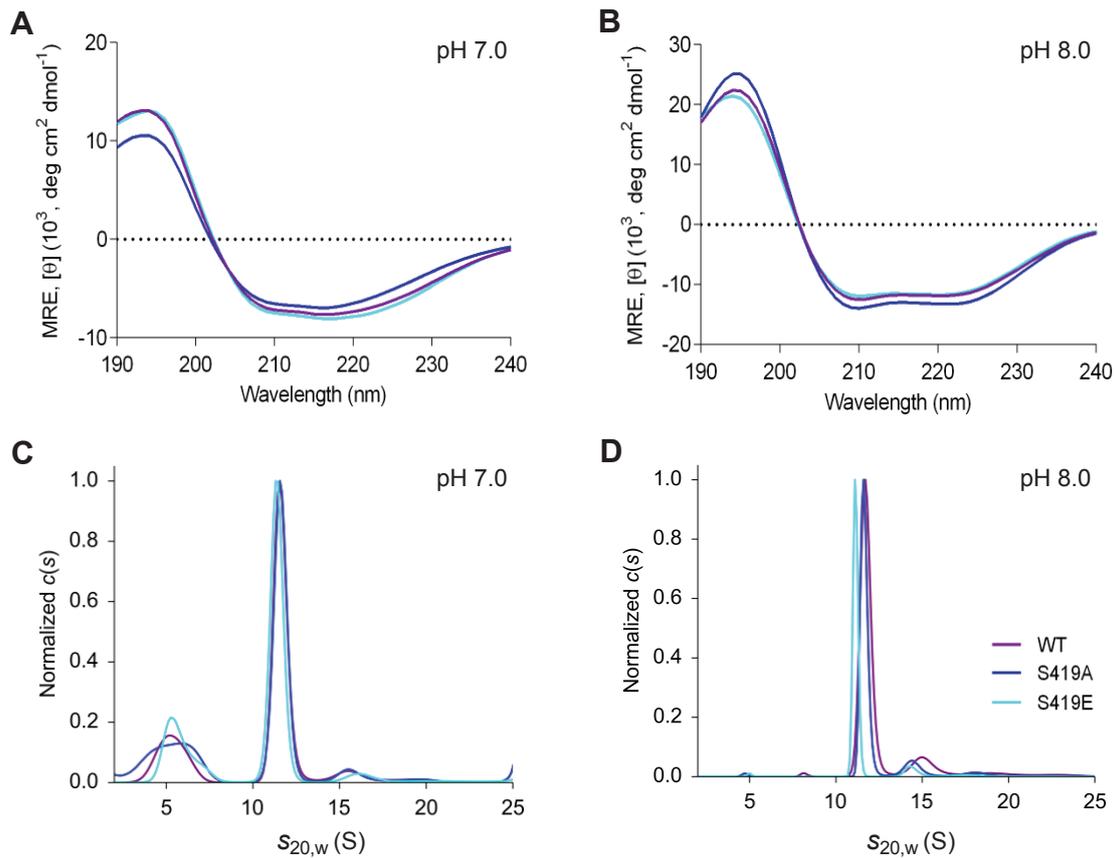
Figure 3. Exploration of the secondary structure organization and quaternary composition of recombinant ZmC₄-NADP-ME versions. A and B: CD spectra of ZmC₄-NADP-ME WT and the two mutated versions obtained at pH 7.0 (A) and 8.0 (B). Ten accumulations each were collected from 190 to 260 nm for 0.16 mg mL⁻¹ enzyme, at 20°C. Each graph is showing the reconstructed curves obtained by applying CONTIN/LL algorithm for data evaluation as provided by the Dichroweb server. **C and D:** Continuous sedimentation coefficient distribution of ZmC₄-NADP-ME WT and the mutated versions at pH 7.0 (C) and 8.0 (D). Data was fitted with the ls-g*(s) model in the software package SEDFIT.
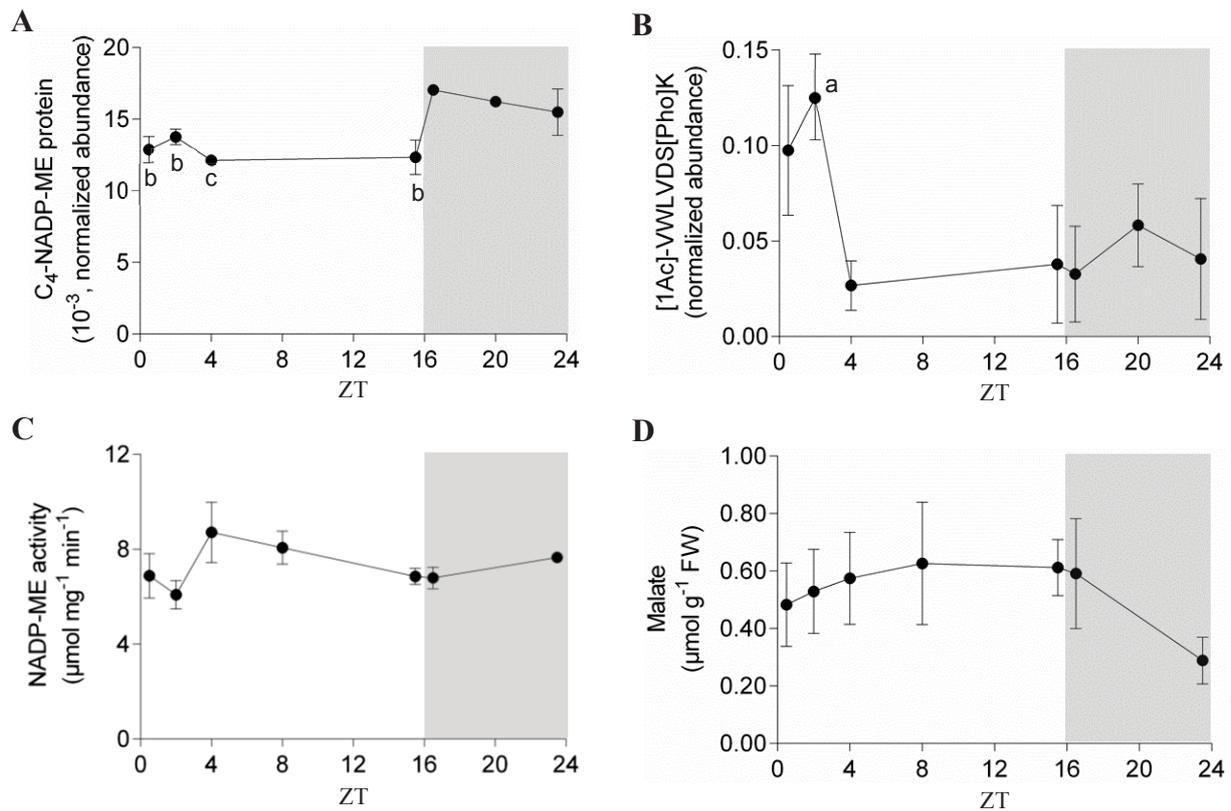
Figure 4. Measurements of different parameters in extracts of maize leaves during a diurnal cycle. A, Quantification of total ZmC$_4$-NADP-ME protein by SWATH-MS. **B**, Quantification of the phosphopeptide VWLVDpSK, corresponding to the phosphorylation of ZmC$_4$-NADP-ME at S419. **C**, Total NADP-ME activity measured with 1.5 - 2.0 μg protein extracts. **D**, Malate concentration assayed by GC-MS analysis. Values represent mean ± standard error of four (A and B) or three (C and D) biological replicates. The dark period is highlighted in grey. Statistical analyses were performed in all cases (A to D) against the first time point in the night (ZT16.5 = 16.5 h). Letters indicate that the value is statistically significant at a, 0.05; b, 0.01; c, 0.001 levels. ZT, Zeitgeber Time.
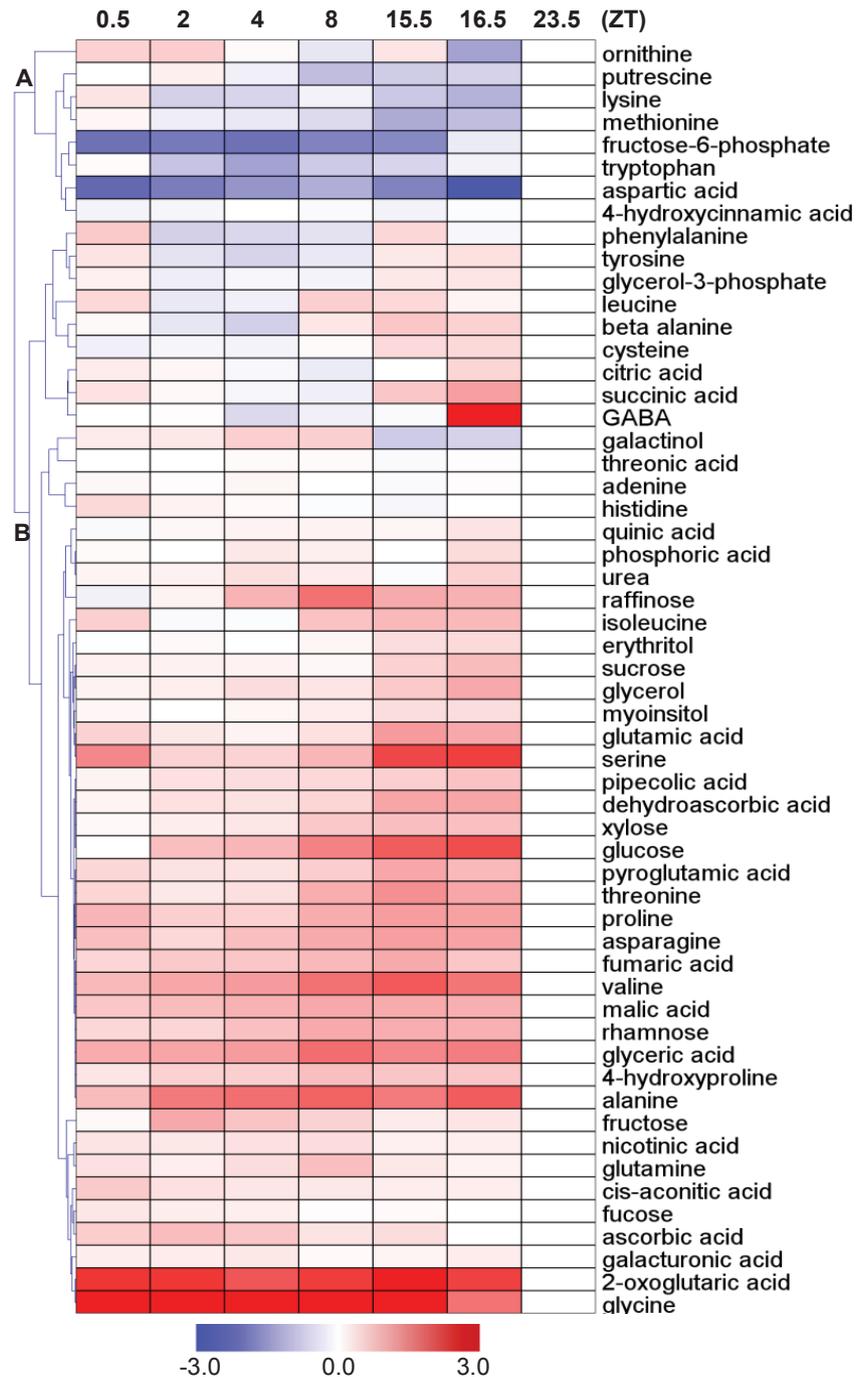
Figure 5. Heat map showing the changes in metabolite relative levels. Metabolites were sorted based on the sum across the time points and then hierarchical cluster was applied. The score, depicted in blue, represents decreased levels, while red color indicates increased levels. The depth of the color corresponds to the magnitude of the change in levels. ZT, Zeitgeber Time. The log2 fold change scale is indicated below the heat map. For details, see Table S2.
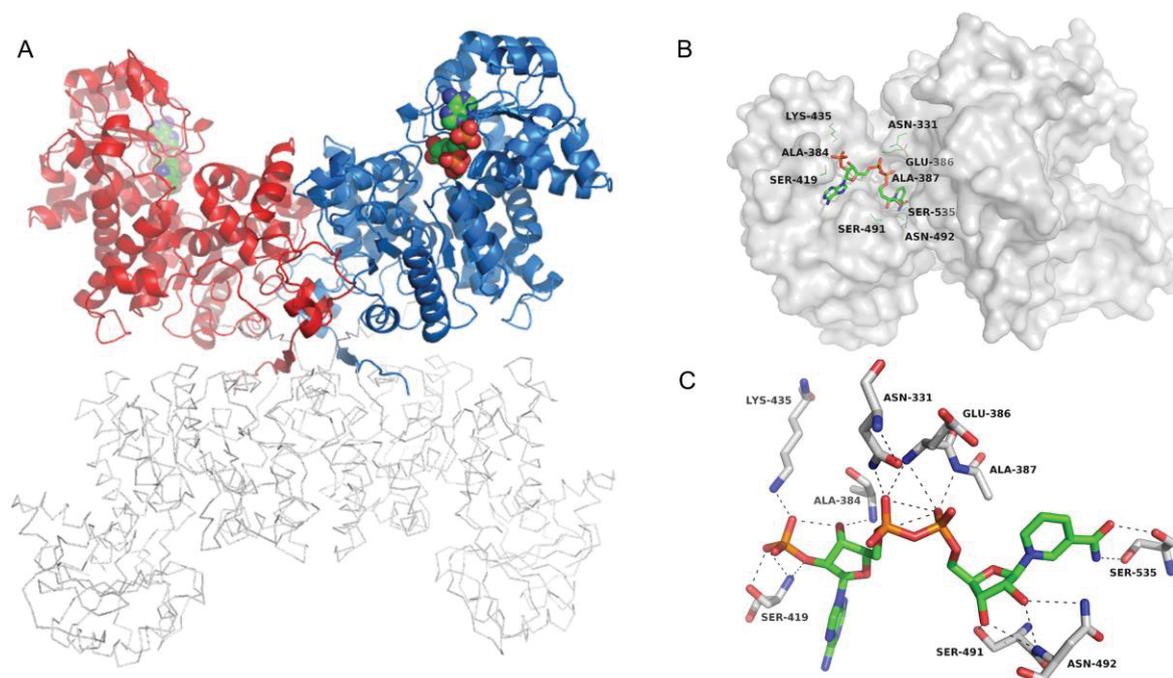
Figure 6. SbC4-NADP-ME crystal structure and amino acids involved in cofactor binding. A, Cartoon and ribbon representation of SbC4-NADP-ME. Monomer A (red) and B (blue) with bound cofactor NADP (colored spheres). Large parts of chain C and D are not well resolved by electron density; therefore, we depicted these chains as gray ribbons only for clarity. **B.** Monomer A from SbC4-NADP-ME with the bound cofactor NADP (represented by sticks). **C.** Residues involved in cofactor binding are depicted as sticks with labels. Hydrogen bonding pattern in the SbC4-NADP-ME cofactor binding site (distances are omitted for clarity and are listed in Table S4).
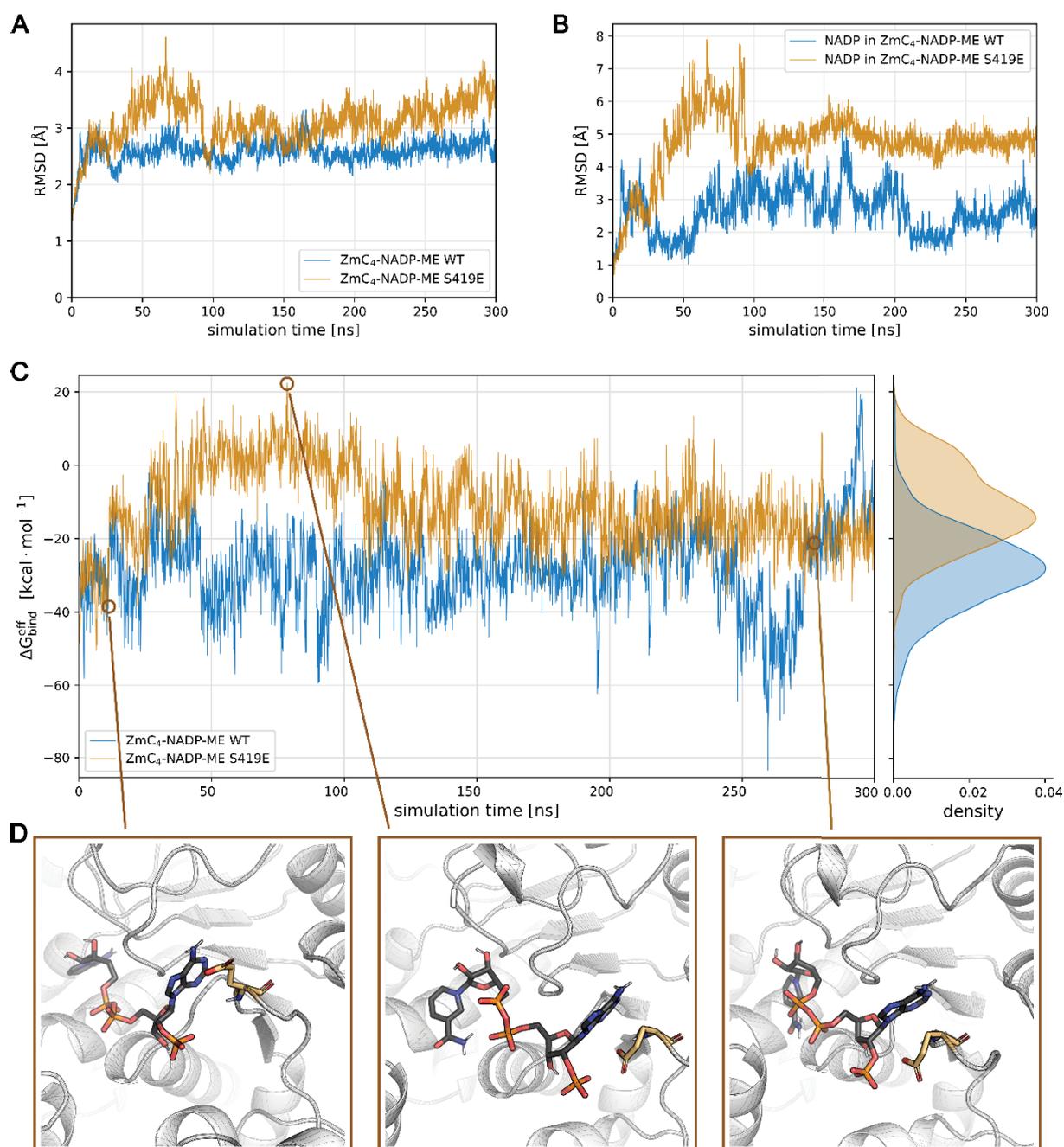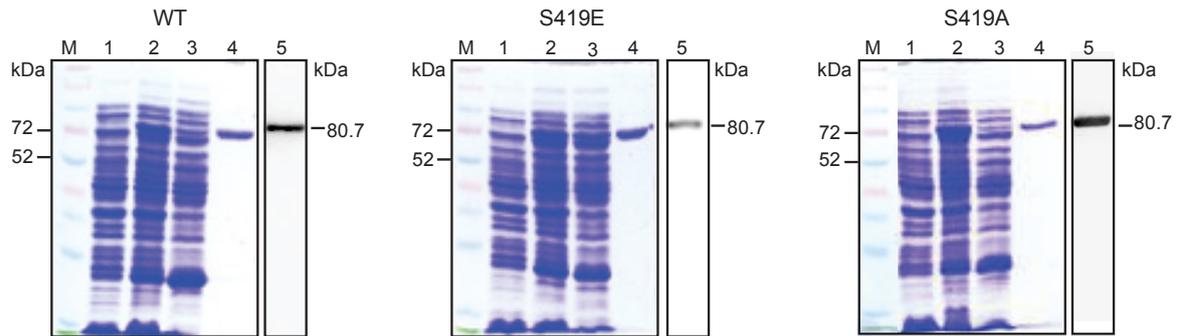
Figure 7. Root mean square deviations and effective binding energies of NADP/ZmC₄-NADP-ME complexes. A. Cα-RMSD of WT ZmC₄-NADP-ME (blue) and the S419E variant (orange) along the MD trajectories. **B.** RMSD of NADP in the binding site after fitting on the enzyme; color code as in panel A. **C.** Time series of the effective binding energies for NADP binding to WT ZmC₄-NADP-ME (blue) and the S419E variant (orange); on the right, the probability distributions of the effective binding energies are shown. **D.** Various binding poses of NADP in the S419E variant at indicated time points in the trajectory; the mutated residue E419 is highlighted in yellow.

Supplementary data

```
                    20            40            60            80
                    |             |             |             |
ZmC4-NADP-ME    MLSTRTAAVAASASPASPWKLGG-RSE--GGASCDGCRTYRNTLRRRAAPAKVRALPPRRVDAVAMVSNAETETEKEQEE 77
SbC4-NADP-ME    .I.A.A.T...-........R..G...--.AG.....S...K.V.....A........K..E...IG.A.....EVVV. 77
SiC4-NADP-ME    ...A..-..SA.....AV.GSK----.G.........ESV....TVR.H.AEQ...Q...VGTA...Q------. 69
ZmnonC4-NADP-ME .I.A...--...--...L..Q.----...G.............-..S.A.E..K.G.A..M.Q.--EA. 70
SbnonC4-NADP-ME .I.A.....-SP.....A..R..G...VG..S..............S..A.....V....E....G.A....K...EQQ 78
Os-NADP-ME      ...A.A..T..A.AASPL..R.EGG.S-GS.SG.TS..E----V.....AVR...AA....E....E.A...KK----. 71

                       100            120           140           160
                       |              |             |             |
ZmC4-NADP-ME    AAAA---------------SEELPVMPWATSVASGYTLLRDPHHNKGLAFTEEERDGHYLRGLLPPAVLSQELQIKKFMN 142
SbC4-NADP-ME    V..----------------N...............R........K..A..........V.........A. 142
SiC4-NADP-ME    GV..ADGGVVDPYAED----........Y................K..A.........V..V.....H 145
ZmnonC4-NADP-ME V..AAGVVVEDHYGEDSAT....IL...F....................M.K...A.H.......VP.......I.H 150
SbnonC4-NADP-ME EVV.GAGGLVEDHYDEDSAA..V.I...F....................K...A.H.......VP.Q..V..I.H 158
Os-NADP-ME      EV..AGGGV------EDMAT..V..T...F...............S.K...A.........V..D..V..I.H 145

                    180            200           220           240
                    |              |             |             |
ZmC4-NADP-ME    TLRQYQTPLQRYIAMMNLQETDERLFYKLLIDNVVELLPFVYTPTVGEACQKYGSIFGRPQGLYVSLKDKGKVLEVLRNW 222
SbC4-NADP-ME    N.....L.I.C.M...............E.......Y..................R.....R 222
SiC4-NADP-ME    N.....L.....M...D...RN.....R.....E...............RQ.........R...R..... 225
ZmnonC4-NADP-ME N.....V.....M...D...RN.........E...V...............RQ.........R.........A 230
SbnonC4-NADP-ME N.....V.....M...D...RN.........E...V...............RQ.........R.........A 238
Os-NADP-ME      N....SV.....M...D...RN.........E...V...............RQ.........D..... 225

                       260           280           300           320
                       |             |             |             |
ZmC4-NADP-ME    PHRNIQVICVTDGERILGLGDLGCQGMGIPVGKLALYTALGGVDPSVCLPITIDVGTNNEKLLNDEFYIGLRQKRATGEE 302
SbC4-NADP-ME    ....V..............................A.........................R... 302
SiC4-NADP-ME    .Q.D........G.......A............R..A..........E..............K. 305
ZmnonC4-NADP-ME .............................S........R..A..................... 310
SbnonC4-NADP-ME ....E......................S........R..A..................... 318
Os-NADP-ME      .E.....V.....................S....R..A.........Q........R....K. 305

                    340           360           380           400
                    |             |             |             |
ZmC4-NADP-ME    YDELIEEFMSAVKQFYGEKVLIQFEDFANHNAFDLLEKYSKSHLVFNDDIQGTASVVLAGLLAALKMVGGTLAEQTYLFL 382
SbC4-NADP-ME    ....M...A...T.......................T..................... 382
SiC4-NADP-ME    .H.......V.I.....................................S..V.......H..... 385
ZmnonC4-NADP-ME .H..M....A...I...............N..A..................V..R....H..... 390
SbnonC4-NADP-ME .H..M....A...I...V...........A.....................V...V.......H..... 398
Os-NADP-ME      .H..M........I...............A.....................SS..V.......H..... 385

                    420           440           460           480
                    |             |             |             |
                                        *
ZmC4-NADP-ME    GAGEAGTGIAELIALEISKQTNAPIEECRKKVWLV[DSKGL]IVDSRKGSLQPFKKPWAHEHEPLKTLYDAVQSIKPTVLIG 462
SbC4-NADP-ME    .................M....K.................S..A..............T........ 462
SiC4-NADP-ME    ...........Q.....K.......Q............S...E..PS..Q.........VS..F........ 465
ZmnonC4-NADP-ME ................S......I............N.V.S.............T........ 470
SbnonC4-NADP-ME ................KT..................N...S.............T........ 478
Os-NADP-ME      ................K.........L..J..N...E...A.........VT..L........ 465

                    500           520           540           560
                    |             |             |             |
ZmC4-NADP-ME    TSGVGRTFTKEIIEAMSSFNERPIIFSLSNPTSHSECTAEQAYTWSQGRSIFASGSPFAPVEYEGKTFVPGQSNNAYIFP 542
SbC4-NADP-ME    ..........V..A.I..............T...AV...........D........ 542
SiC4-NADP-ME    .......A...VV..A......V............N.TK..AV.....N...D.......A...... 545
ZmnonC4-NADP-ME ..........VV..A..............E..N.T...V........D...D.V.......... 550
SbnonC4-NADP-ME ..........VV..A..............E..N.T..A........D...D..I....A...... 558
Os-NADP-ME      .....K....V...A......V........E..N...AV.....D....N..IH...... 545

                    580           600           620           640
                    |             |             |             |
ZmC4-NADP-ME    GLGLGLVISGAVRVHEDMLLAASKALADQATQDNFEKGSIFPPFTSIRKISAHIAAAVAAKAYELGLATRLPPPSDLVKY 622
SbC4-NADP-ME    .....................A......EE..VT.......N......Y..................K...A. 622
SiC4-NADP-ME    ......IMA..T.........E.......N.E...........S.................A.R....Q. 625
ZmnonC4-NADP-ME .F...........I.......E..A...E..D........N......R...............R..... 630
SbnonC4-NADP-ME .F...........I.......E..G...E..D........N......R...............R..... 638
Os-NADP-ME      .F...V..............ET.......E.............N......R...S.........Q.R..E.. 625


ZmC4-NADP-ME    AENCMYTPVYRNYR 636
SbC4-NADP-ME    ..S...S......Q 636
SiC4-NADP-ME    ..T...N.I..... 639
ZmnonC4-NADP-ME .QS........... 644
SbnonC4-NADP-ME ..S........... 652
Os-NADP-ME      ..S........S.. 639
```

Supplemental Figure 1. Alignment of plastidic C₄- and nonC₄-NADP-ME sequences in Poaceae showing conservation of Ser419 and the BARK substrate motif (DSKGL). We obtained the amino acid sequences of the plastidic C₄- and nonC₄-NADP-ME isoforms of *Z. mays* cultivar Ensembl-18 (Sequence IDs: GRMZM2G085019_T01, GRMZM2G122479_P01) and *S. bicolor* cultivar v3.1.1 (Sequence IDs: Sobic.003G036200.1, Sobic.009G108700.1) from the Phytozome 11 database. We used the maize C₄ and nonC₄-NADP-ME sequences as BLAST queries to identify the corresponding orthologs in the *Setaria italica* cultivar v2.2 proteome in the same database (Sequence ID: Seita.5G134300.1). We also used the maize nonC₄-NADP-ME sequence as a BLAST query to identify the unique plastidial NADP-ME in the *Oryza sativa* cultivar v7_JGI proteome (LOC_Os01g09320.1). All

sequences were aligned using PRANK. The asterisks (*) denote the position of Ser419. The green box indicates the BARK substrate motif, which includes the Ser419, among plastidic photosynthetic and non-photosynthetic NADP-ME isoforms within the Poaceae. Zm: *Zea mays*; Sb: *Sorghum bicolor*; Si: *Setaria italica*; Os: *Oryza sativa*.

Supplemental Figure 2. SDS-PAGE stained with Coomassie-Blue (lines 1-4) and analysed by immunoblot (line 5) of protein fractions during the isolation of recombinant ZmC4-NADP-ME variants. M, molecular weight markers (Spectra Multicolor Broad range Protein Ladder (Thermo Fisher Scientific); 1, 20 µg of non-induced cell culture lysate; 2, 20 µg of cell culture lysate 16 h after induction of protein production; 3, 20 µg of crude extract cleared by centrifugation; 4 and 5, 1.5 µg of affinity purified protein; The size of the recombinant proteins (80.7 kDa; mature ZmC4-NADP-ME 63.4 kDa plus 17.3 kDa encoded by the expression vector) is indicated on the left. Anti-His-HRP conjugate antibodies were used for the immunoblots.

Supplemental Figure 3. Cartoon and surface representation of SbC$_4$-NADP-ME chain B.
The bound cofactor NADP and the phosphorylation site Ser419, located on the surface of the
enzyme, are highlighted as sticks.

Supplemental Table 1. Secondary structure contents of ZmC$_4$-NADP-ME WT, S419A and S419E estimated from the CD-spectra obtained at pH of 7.0 and 8.0. The Dichroweb server (http://dichroweb.cryst.bbk.ac.uk) was used to calculate protein secondary structure contents from experimentally acquired data. The secondary structures are presented as a sum of regular α-helix, distorted α-helix, regular β-strand, distorted β-strand, turns and disordered conformations in case of each analysed protein. The normalized root-mean-square deviations, a measure for the quality of the data evaluation expressed as the difference between the original CD spectrum and its reconstruction, were always lower than 0.1.

		Regular α-helix	Distorted α-helix	Regular β-strand	Distorted β-strand	Turns	Disordered	Total
pH 7.0	WT	0.136	0.100	0.185	0.096	0.187	0.296	1
	S419A	0.108	0.099	0.190	0.100	0.198	0.304	0.999
	S419E	0.143	0.115	0.165	0.093	0.195	0.290	1.001
pH 8.0	WT	0.228	0.143	0.128	0.067	0.168	0.265	0.999
	S419A	0.258	0.175	0.079	0.059	0.181	0.248	1
	S419E	0.221	0.140	0.127	0.073	0.175	0.264	1

Supplemental Table 2. Changes in metabolites in maize leaves. Material was harvested at ZT0.5, 2, 4, 8, 15.5, 16.5, and 23.5 (the last two time points represent the beginning and end of night period, respectively). Values presented are mean ± standard error of three replicates. Bold letters indicate significant differences by Student´s t test ($P < 0.05$) compared to the ZT23.5 time point.

	0.5 h	2 h	4 h	8 h	15.5 h	16.5 h	23.5 h
cis-aconitic acid	**4.4 ±0**	3.6 ±0.3	3.4 ±0.3	3.4 ±0.3	3.3 ±0.3	3.3 ±0.4	2.8 ±0.2
adenine	1.3 ±0.1	1.3 ±0.1	1.3 ±0.1	1.2 ±0.1	1.2 ±0.1	1.3 ±0	1.2 ±0.1
alanine	1596.8 ±382.4	2720.9 ±531	**3007 ±474.6**	3333 ±713.5	**2703.7 ±370**	3567.8 ±647.7	929.4 ±197.3
beta-alanine	6.8 ±1.5	5 ±1.3	4 ±1	7.9 ±1.2	10.2 ±1.4	9.4 ±1.2	6.4 ±1.2
ascorbic acid	19 ±5.8	21.6 ±7.9	20.1 ±7.3	15.8 ±7.2	16.6 ±7.4	12.8 ±2.9	12.6 ±2.6
asparagine	5.5 ±1.1	4.5 ±0.9	5.5 ±0.9	6.5 ±1.3	**7.1 ±0.3**	6.8 ±1.4	3.3 ±0.5
aspartic acid	**123.4 ±23.6**	**156 ±17.8**	**199.9 ±34.6**	252.3 ±81.5	**166.9 ±28.1**	**78.3 ±10.2**	561.3 ±68.9
4-hydroxycinnamic acid	3.7 ±0.5	3.8 ±0.3	4.2 ±0.4	4 ±0.8	3.7 ±0.3	4.1 ±0.4	4.2 ±0.3
citric acid	215.9 ±33.7	198 ±25.7	174.1 ±25	152.4 ±22.7	185.3 ±12.4	258.9 ±40.9	183.6 ±19.9
cysteine	2.8 ±0.4	2.9 ±0.2	2.9 ±0.3	3.4 ±0.9	4.4 ±1.1	4.4 ±0.6	3.2 ±0.3
dehydroascorbic acid	13.5 ±0.6	15.9 ±1.4	15.4 ±1.3	**17.4 ±1.1**	**25.2 ±0.7**	**25.2 ±0.4**	12.3 ±0.7
erythritol	7.4 ±0.2	8 ±0.3	7.6 ±0.1	8.1 ±0.3	**9.8 ±0.2**	**10.2 ±0.2**	7.5 ±0.3
fructose	181.6 ±19.6	343.4 ±72	276.1 ±51.9	244.3 ±41.1	203.6 ±20	212.5 ±11.9	170.9 ±16.4
fructose-6-phosphate	**3.6 ±0.3**	**3.9 ±0.3**	**3.6 ±0.5**	**4.2 ±0.7**	**4.5 ±0.7**	11.8 ±1.8	14.3 ±1.8
fucose	23.9 ±1.1	22.7 ±1.8	22.6 ±1	20 ±0.4	20.6 ±2	19.7 ±1.6	19.5 ±2.3
fumaric acid	10.9 ±1.6	12 ±1.7	12.4 ±1.5	13.7 ±2.1	**15.3 ±1.6**	12.4 ±1.4	7.7 ±0.9
GABA	0.6 ±0.1	0.6 ±0.1	0.4 ±0.1	0.5 ±0	0.5 ±0.1	8.5 ±2.1	0.6 ±0.1
galactinol	97.3 ±3.6	99.2 ±11.8	**121.6 ±3.1**	120.4 ±12	49 ±7.7	52.9 ±7	81.9 ±9.2
galacturonic acid	0.7 ±0	0.7 ±0	0.7 ±0	0.6 ±0	0.7 ±0	0.7 ±0	0.6 ±0
glucose	61 ±12.1	101.7 ±26.5	107.9 ±27.7	166.4 ±44	**232.6 ±27.6**	**264.7 ±40.3**	60.1 ±8.2
glutamic acid	863.7 ±105.4	726.9 ±82	670.8 ±77.9	771.8 ±174.5	**1340.7 ±160.2**	1211.8 ±210.3	601 ±78.6
glutamine	49.2 ±3.5	44.2 ±4.8	50.7 ±6.4	64.7 ±17	46.3 ±5.3	43.1 ±5.6	38.6 ±3.2
2-oxoglutaric acid	**39.5 ±0.5**	**39.4 ±5.5**	**28.8 ±1.3**	**36.6 ±5.2**	**62.3 ±10.9**	**34.8 ±6.4**	7.1 ±1
glyceric acid	21.9 ±3.4	**23.1 ±2.5**	**25 ±1.2**	**36.8 ±4**	**29.7 ±2.4**	**31.9 ±4.5**	11.2 ±1.6
glycerol	25.9 ±0.6	26.9 ±4.5	30.5 ±4.3	28.7 ±3.2	35.8 ±4	**46.3 ±3.7**	23.2 ±1.1
glycerol-3-phosphate	42.1 ±4.4	31.4 ±6.2	34.4 ±7.3	33.4 ±2.9	44.9 ±8.3	45.9 ±8.4	37.1 ±6.5
glycine	**584.7 ±95.1**	**361.2 ±37.1**	**384.7 ±55.1**	**425.2 ±75.2**	**734.6 ±86.1**	139.8 ±41.6	44.6 ±7.7
histidine	7.5 ±0.5	6.2 ±0.5	5.8 ±0.3	5.4 ±0.6	5.2 ±0.6	5.5 ±0.6	5.5 ±0.7
myoinositol	335.8 ±10.8	311.8 ±11	336.6 ±17.1	365.9 ±18.1	**406.3 ±16.8**	**409.2 ±17.2**	310.4 ±10.8
isoleucine	18 ±3.1	11.7 ±2	11.9 ±1.7	20 ±1.9	21.2 ±1.5	21.2 ±0.9	12.2 ±2.2
leucine	29.1 ±3.9	17.2 ±2.9	18.5 ±1.6	31.2 ±2.5	29.3 ±2.2	23.2 ±1.8	21.2 ±3.8
lysine	10.8 ±0.1	5.5 ±0.6	5.7 ±0.3	7.8 ±0.9	5 ±0.4	4.1 ±0.8	8.7 ±1.1
malic acid	486.2 ±94	528.4 ±94.4	570.1 ±103.6	617.6 ±137.5	604.4 ±63.2	584.9 ±123.7	308.5 ±52.7
methionine	2 ±0.3	1.6 ±0.2	1.5 ±0.3	1.3 ±0.2	**0.8 ±0.1**	1 ±0.2	1.8 ±0.1
nicotinic acid	1.6 ±0.2	1.6 ±0.2	1.7 ±0.2	1.7 ±0.2	1.5 ±0.1	1.5 ±0.3	1.3 ±0.1
ornithine	2.6 ±0.6	2.7 ±0.7	1.9 ±0.5	1.4 ±0.5	2.2 ±0.4	0.7 ±0.1	1.8 ±0.5
phenylalanine	9.2 ±1.1	3.7 ±0.4	4 ±0.5	4.5 ±0.2	8.3 ±0.2	5.6 ±0.4	5.9 ±0.6
phosphoric acid	494.3 ±66.9	482.1 ±63.4	569.3 ±112.7	541.6 ±88.1	475.3 ±85.4	641 ±77.2	475.4 ±89.9
pipecolic acid	1.5 ±0.1	1.7 ±0.2	1.7 ±0.3	1.8 ±0.1	**2 ±0.1**	2.2 ±0.4	1.3 ±0.1
4-hydroxyproline	1 ±0.2	1.1 ±0.1	1.1 ±0.2	1.3 ±0.3	1.2 ±0.2	1.2 ±0.3	0.8 ±0.1
proline	20.6 ±7.9	16.8 ±4	16.5 ±4.9	22.1 ±6.4	25.1 ±6.7	24.3 ±8.2	11.4 ±2.3
putrescine	103.4 ±15	116.3 ±19	88.8 ±13.2	53.1 ±2.4	61.9 ±5.5	65.6 ±7	102 ±14
pyroglutamic acid	268.7 ±29.7	243.5 ±28.4	244.3 ±30.1	287.2 ±62	**391.4 ±45.5**	336.6 ±57.2	192.6 ±21.5
quinic acid	483.2 ±75.5	538.6 ±100.9	563.6 ±47.5	557.9 ±59.5	546.1 ±80.9	627.2 ±101.2	503.1 ±44
raffinose	90.5 ±11.9	113.5 ±5.2	188.8 ±13.4	**323.9 ±36.3**	201.8 ±28.7	191 ±28	102.3 ±17.6
rhamnose	**2.8 ±0.2**	2.9 ±0.4	**3.4 ±0.1**	**4 ±0.2**	3.9 ±0.5	3.8 ±0.6	2 ±0.1
serine	**414.3 ±53.7**	222.5 ±30.6	222.9 ±36.3	279.9 ±65.8	**726.7 ±98.4**	**799.2 ±148.8**	156.7 ±26.7
succinic acid	26.2 ±3.8	22.1 ±2.6	19.1 ±2.9	17.8 ±1.4	33 ±2.2	**45.2 ±2.7**	20.8 ±2.6
sucrose	1183.5 ±84.5	1175 ±267.8	1166.4 ±174.4	1117.1 ±175.2	1516.2 ±159.8	**1793.9 ±157**	1043.1 ±27.7
threonic acid	82.3 ±14.7	82.1 ±13.2	85.6 ±10	85.2 ±10.8	78.6 ±8	84.4 ±11.5	82 ±10.4
threonine	6.6 ±0.7	5.7 ±1	6.1 ±0.9	9 ±1.8	11.5 ±1.8	9.5 ±1.1	4.7 ±0.6
tryptophan	10.5 ±1.7	5.7 ±0.7	**4.1 ±0.3**	6 ±1.6	6.7 ±1.2	9.1 ±2.3	10.2 ±1.2
tyrosine	69 ±8.7	42.4 ±4.8	36.1 ±3.6	45 ±2.4	66.6 ±2.9	71.3 ±2.2	55.4 ±5.7
urea	9.5 ±1.5	9.7 ±1.4	11.2 ±2.6	10.1 ±1.7	8.5 ±1.1	12.5 ±1.7	8.6 ±1.3
valine	80 ±8.8	92.6 ±11.4	**101.3 ±8.9**	**143.5 ±18.9**	**178.7 ±23.6**	**137.7 ±12**	45.2 ±4.9
xylose	2.1 ±0.1	2.3 ±0.2	2.4 ±0.1	**3.1 ±0.1**	**3.4 ±0.1**	**3.3 ±0.1**	2 ±0.1

Supplemental Table 3. Protein structure data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

	SbC$_4$-NADP-ME pdb-code: 6C7N
Wavelength [Å]	1.5418Å
Resolution range [Å]	29.28 - 2.00 (2.03 - 2.00)
Space group	C 1 2 1
Unit cell [Å; °]	208.10, 64.26, 202.72 90 94 90
Total reflections	665459 (31790)
Unique reflections	181154 (8879)
Multiplicity	3.7 (3.6)
Completeness	1.00 (1.00)
Mean I/sigma(I)	9.1 (2.1)
Wilson B-factor [Å^2]	18.2
R-merge	0.085 (0.585)
R-meas	0.100 (0.690)
CC1/2	0.997 (0.733)
CC*	1 (0.946)
Reflections used in refinement	181074 (17968)
Reflections used for R-free	1819 (164)
R-work	0.180 (0.211)
R-free	0.210 (0.252)
CC(work)	0.950 (0.871)
CC(free)	0.941 (0.764)
Number of non-H atoms	18830
macromolecules	15643
ligands	297
Protein residues	2030
RMS(bonds)	0.010
RMS(angles)	0.99
Ramachandran favored (%)	98.20
Ramachandran allowed (%)	1.8
Ramachandran outliers (%)	0
Rotamer outliers (%)	2.0
Clashscore	1.64
Average B-factor [Å^2]	33.42
macromolecules	32.46
ligands	32.46
solvent	40.90

Supplemental Table 4. Hydrogen bonds with and distances to the cofactor NADP in monomer A of SbC$_4$-NADP-ME.

Atom	Residue	Atom	Distance [Å]
ND2	Asn331	O1A	2.75
N	Ala384	O3B	2.92
N	Glu386	O1A	3.08
N	Ala387	O2N	2.82
N	Ser419	O1X	3.19
N	Ser419	O2B	3.26
OG	Ser419	O1X	2.60
NZ	Lys435	O3X	2.73
OG	Ser491	O3D	3.32
N	Asn492	O3D	3.29
OG	Ser535	N7N	2.66
OD1	Asn537	O7N	2.93

6 Discussion

6.1 Molecular adaptations of NADP-malic enzyme for its function in C₄ photosynthesis in grasses

The maize chloroplastic *C₄-NADP-ME* gene has evolved through duplication of the gene encoding an ancestral plastidic nonC₄-NADP-ME isoform (Tausta et al., 2002). In order to fulfill the C₄ function, changes in the *cis* regulatory sequence of the newly evolved gene resulted in the characteristic BS cell specificity and regulation by light (Tausta et al., 2002). In addition, since the molecular surrounding within the chloroplasts of the C₄ BS cells differs from the conditions within the mesophyll cells (Turkan et al., 2018), changes in the coding sequence underpinning the changes of the enzymatic and regulatory properties of the newly evolved isoform were also required (Christin et al., 2009).

In manuscript I, in order to identify amino acids responsible for the unique C₄ properties of C₄-NADP-ME, we obtained crystal structures of the maize C₄-NADP-ME in its ligand-free state and the sorghum C₄-NADP-ME with bound cofactor. This approach and phylogenetic analyses allowed us to identify four strictly differentially conserved amino acids between the C₄- and nonC₄-NADP-ME isoforms from grasses. The contribution of these residues to the C₄ molecular characteristics was examined in detail by a series of kinetic and structural studies.

6.1.1 Molecular adaptations of the C₄-NADP-ME required for its tetrameric assembly

From the four strictly conserved amino acid residues identified in our study only Phe140 lies within a region shown to be involved in the oligomerization of the protein (Detarsio et al., 2007). After changing this amino acid to the one present in the nonC₄ isosforms, the mutagenized decarboxylase showed impaired quaternary structure (manuscript I, Fig. 5a and 5c). The crystal structure of ZmC₄-NADP-ME indicated that F140 is located close to the dimer interface and might be important for the stability of the helical element αA3 (manuscript I, Fig. 6a).

However, the facts that the F140I mutation introduced to the ZmC₄-NADP-ME did not lead to the homodimeric assembly and the reverse mutation introduced to the ZmnonC₄-NADP-ME did not result in the tetrameric oligomerization state (manuscript I, Fig. 4d), indicate that additional amino acid residues must be involved in conferring the homotetrameric assembly of the photosynthetic isoform. This is in agreement with findings of Detarsio and colleagues (Detarsio et al., 2007), who reported that the chimeric nonC₄-NADP-ME decarboxylase

containing the amino acids 102 till 247 of the C4 isoform assembled as a mixture of dimers and tetramers (Detarsio et al., 2007).

In this work we found that at pH 7.0 the WT ZmC4-NADP-ME is also present in a lower oligomerization state (manuscript I, Fig. 5d). An equilibrium between the homotetramers and homodimers at pH 7.0 has also been described for the C4-NADP-ME from sugarcane (Iglesias and Andreo, 1990). Moreover, the enzyme in the lower oligomerization state is inactive as in the case of the sorghum C4-NADP-ME (Saigo et al., 2013). These structural changes can be of physiological significance, since the activity of the C4-NADP-ME decarboxylase should be maximized during the day (to ensure the rapid CO_2 supply for RubisCO) and reduced in the night (Drincovich and Andreo, 1994), when the C4 pathway is less or not active (Buchanan, 1980). In agreement with this, ZmC4-NADP-ME is assembled solely as active homotetramer at pH 8.0 (manuscript I, Fig. 5a and 5b), which is the approximate pH value in the chloroplasts under illumination (Heldt et al., 1973; Werdan et al., 1975). In the night the pH in the chloroplasts decreases to 7.0 or 7.5 (Heldt et al., 1973; Werdan et al., 1975), which would provoke conformational changes of the ZmC4-NADP-ME, so that a fraction of inactive dimers or monomers is formed and the activity of the decarboxylase is therefore reduced.

Taken altogether, the F140 residue represents the first step towards identification of the amino acid residues governing the assembly and stability of the ZmC4-NADP-ME's in its active homotetrameric state. For the complete elucidation of the structural behaviour of the photosynthetic NADP-ME isoform, further site-directed mutagenesis studies are required. The most promising candidates for these studies are amino acid residues localized at the dimer interface (manuscript I, Suppl. Table 2).

6.1.2 Molecular adaptations of the C4-NADP-ME required for its kinetic and regulatory properties

Another particularity of the C4-NADP-ME isoform from maize (Detarsio et al., 2007; Saigo et al., 2013), sorghum (Saigo et al., 2013) and sugarcane (Iglesias and Andreo, 1990) is its allosteric inhibition by malate at pH 7.0-7.5. This might represent a regulatory mechanism, since it might contribute to decarboxylase activity reduction during the night.

In our work we show that Glu339 is required for malate inhibition, since the exchange of Glu339 to Ala, which is found in the nonC4-isoforms in grasses, was sufficient for a complete abolishment of this regulatory feature (manuscript I, Fig. 3e). The reverse amino acid exchange in the ZmnonC4-NADP-ME resulted in the acquisition of malate inhibition (manuscript I, Fig. 4c). The crystal structures of the two C4-NADP-MEs obtained in this study enabled to

determine the exact positioning of the Glu339 within the homotetramer. This amino acid lies far from the active center (manuscript I, Fig. 6d), on the surface of helix αB4 and its side chain points out to the surrounding solvent. We propose that this is the location of the allosteric binding site. However, for the ultimate confirmation of this conclusion, a crystal structure of the C$_4$-NADP-ME in complex with bound malate at this site is required, similar to the studies performed with the photosynthetic PEPC isoform from *Flaveria trinerva* (Paulus et al., 2013). In this case the allosteric site for malate/aspartate was unequivocally identified by X-ray crystallization studies and further confirmed by site-directed mutagenesis (Paulus et al., 2013). Nevertheless, our results confirm the results of Detarsio and colleagues, who showed that the region of ZmC$_4$-NADP-ME from amino acid residue 248 to the C-terminus is critical for the allosteric malate inhibition in a pH-dependent manner (Detarsio et al., 2007).

The C$_4$-NADP-ME isoform differs from its nonC$_4$ counterpart not only by its structural and regulatory properties, but also with regards to the main kinetic properties (Saigo et al., 2004; Alvarez et al., 2013; Saigo et al., 2013). These differences include the 2 to 5 times higher catalytic efficiency, twofold higher affinity for malate and 2 to 9 times higher affinity for NADP compared to the nonC$_4$-NADP-ME isoforms. In our study two of the four strictly differentially conserved amino acid residues identified, Gln503 and Leu544, when exchanged in the ZmC$_4$-NADP-ME to those found in the nonC$_4$ isoforms, resulted in enzymes with reduced affinity to malate. The K$_m$ values obtained in case of the both mutated proteins were roughly twofold higher than that of the WT (manuscript I, Fig. 3f), being thus closer to the corresponding nonC$_4$ values. These results suggest that during the neo-functionalization of the ancestral chloroplastic NADP-ME, the charged Glu503 and aromatic Phe544 were changed to polar uncharged Gln503 and aliphatic Leu544 respectively in order to provide the resulting C$_4$ version with an increased affinity for malate.

6.2 Involvement of post-translational modifications in the modulation of ZmC$_4$-NADP-ME activity

Processes involved in the regulation of C$_4$ photosynthesis are complex and operate at multiple levels: epigenetic, transcriptional, post-transcriptional, translational and post-translational (Reeves et al., 2017). Whereas the first four regulatory mechanisms are quite extensively studied (Wang et al., 2011; Reeves et al., 2017), less information is available concerning the involvement of PTMs in governing the activity of C$_4$ proteins. A strategy to analyse protein phosphorylation is the use of high-throughput MS-based phosphoproteomics together with specific phosphopeptide enrichment techniques (Mann et al., 2002).

In manuscript II, a phosphorylation site at Ser419 was detected in the maize C_4-NADP-ME isoform by LC-MS, and a series of experiments were performed to analyse the significance of the discovered phosphorylation site.

6.2.1 Influence of Ser419 phosphorylation on the activity of ZmC₄-NADP-ME

In order to answer the questions of whether and how the identified phosphorylation event influences the enzymatic activity, mutated ZmC_4-NADP-ME variants were generated and expressed in a heterologous system. In order to mimic the phosphorylated state of the protein, the Ser419 was changed to glutamic acid (S419E); as a not phosphorylated control the serine residue was changed to alanine (S419A). This is a widely used technique for establishing the influence of phosphorylation on the enzymatic and structural properties of a protein (Chen and Cole, 2015).

In manuscript II we show that the S419E variant has a highly decreased affinity to malate and to the cofactor $NADP^+$, and a lower total catalytic activity (manuscript II, Fig. 2A). At pH 8.0 the S419E exhibited very low residual activity, indicating that Ser419 phosphorylation negatively influences the activity *in vivo* (manuscript II, Fig. 2A and 2B). In the case of S419A, a much less dramatic decrease in the affinity to both the substrate and the cofactor was measured (manuscript II, Fig. 2B).

Structural analyses performed with all three protein variants indicated that the introduced mutations did not lead to any significant changes of the secondary (manuscript II, Fig. 3A and 3B) and quaternary (manuscript II, Fig. 3C and 3D) protein organization. This indicates that the observed differences in the enzymatic behaviour are not due to structural changes of the protein. Moreover, these results imply that the additional negative charge introduced by the Ser419 phosphorylation *in vivo* may rather have an effect on the local protein surrounding and is not leading to major changes in protein conformation.

To clarify the molecular mechanisms of the influence of phosphorylation of the Ser419, we have solved the protein structure of the orthologous C_4-NADP-ME isoform from sorghum with bound $NADP^+$ (manuscript II, Fig. 6). Sorghum C_4-NADP-ME shares a similar overall protein conformation with the ZmC_4-NADP-ME (manuscript I). According to the protein structure, Ser419 is directly participating in the binding of the cofactor via hydrogen bond formation (manuscript II, Fig. 6B). This agrees with the results of kinetic measurements performed with $NADP^+$ and heterologous S419E ZmC_4-NADP-ME. The lower activity of the S419A variant can be explained by the abolishment of the hydrogen bond, which cannot be formed when an alanine residue is placed at this position.

The affected serine residue is localized on the solvent-exposed surface of the decarboxylase, which makes it potentially accessible to a kinase and phosphatase (manuscript II, Suppl. Fig. 3). Molecular dynamic simulation experiments performed with the WT and S419E ZmC4-NADP-MEs predicted that $NADP^+$ binding would be less favorable for the phosphomimetic form than for the WT enzyme (manuscript II, Fig. 7A and 7B).

Altogether, the *in vivo* phosphorylation of the ZmC4-NADP-ME at Ser419 decreases the enzymatic performance by creating adverse conditions for $NADP^+$ binding without having a major structural impact on the overall secondary or quaternary protein conformation.

6.2.2 Physiological significance of the phosphorylation at Ser419

In order to assess the physiological significance of the identified phosphorylation site within ZmC4-NADP-ME, we quantified the fraction of phosphorylated protein by sequential window acquisition of all theoretical mass spectra (SWATH)-MS (Ludwig et al., 2018) at several time points during the diurnal cycle. The highest amount of phosphorylation represented about 12% of the total photosynthetic NADP-ME pool and was observed at two hours into light (manuscript II, Fig. 4B). The finding that the maximal detected levels of phosphorylation represented less than a quarter of the total peptide pool implies that phosphorylation at Ser419 residue most probably represents a fine-tuning mechanism for ZmC4-NADP-ME activity regulation.

In maize, the phosphorylation of PEPC, PEP-CK and PPDK plays a pivotal regulatory role in adapting enzymatic activity during the day-night cycle. For instance, Chen and colleagues showed that 30% of PPDK was still phosphorylated at Thr527 at time point of lowest phosphorylation levels (in the middle of the day), while the highest levels of phosphorylation reached 90% during the night (Chen et al., 2014). Similarly, for maize PEPC antibodies specific against peptides harboring phosphorylated Ser15 showed very high amounts of phosphorylation at mid-morning and only residual amounts of phosphorylation at night (Ueno et al., 2000).

To determine whether a correlation between the observed phosphorylation levels and the ZmC4-NADP-ME's enzymatic activity exists, activity measurements were performed with total protein extracts from maize leaves harvested at various time points in the diurnal cycle (manuscript II, Fig. 4C). It is reasonable to assume that the results of the decarboxylation activity measurements can mainly be associated with the C4-NADP-ME isoform activity as this is the isoform with the highest expression in leaf blades (Alvarez et al., 2013) and a good correlation between transcript and protein abundance in maize has been previously

demonstrated (Li et al., 2010). We found only minor changes in the NADP-ME decarboxylation activity during the diurnal cycle. This is consistent with only a small fraction of the ZmC_4-NADP-ME cellular pool being phosphorylated.

In addition, determination of the malate content in maize leaves also showed no drastic changes during the diurnal cycle (manuscript II, Fig. 4D). This is different from C_3 plant Arabidopsis, in which malate levels steadily increase under illumination when assayed under long day conditions; malate levels reach a summit after about 12h of illumination and quickly drop off in darkness (Fahnenstich et al., 2007). However, we have only measured the total malate levels in maize leaves. It has been shown that malate occurs in both photosynthetically active and inactive pools (Arrivault et al., 2017), thus more elaborate experiments will be needed in order to determine exact diurnal changes in the photosynthetic malate content within maize BS cells.

6.2.3 Different regulatory mechanisms modulate ZmC_4-NADP-ME activity within a diurnal cycle

For efficient functioning of the C_4 pathway, the balance between the rate at which carbon is transported into the BS cells and its subsequent utilization within the Calvin-Benson cycle is critical (Sage, 2014; Arrivault et al., 2017). If the CO_2 concentrating mechanism is more active than the C_3 cycle, it is inevitable that CO_2 will overaccumulate in the BS cells and leak back to the M cells. This results in energy waste. Alternatively, if the C_4 cycle is slower than the C_3 cycle, the CO_2 pool in BS cells is exhausted by RubisCO and reaches critical values at which photorespiration and substrate limitation become important (Sage, 2014; Arrivault et al., 2017). This also mitigates the benefits of the C_4 photosynthesis. Since the C_4-NADP-ME isoform is the major decarboxylase in maize BS cells, the regulation of its activity might be necessary for keeping the C_4 cycle in balance.

By exhibiting the highest enzymatic activity under illumination (Drincovich and Andreo, 1994), the ZmC_4-NADP-ME can efficiently supply CO_2 for RubisCO and help maintain the overall high activity of the C_4 cycle. This high activity is enabled by the formation of an active homotetrameric enzyme at the stromal pH value of illuminated chloroplasts. During the night the stromal pH decreases to 7.0-7.5 at which ZmC_4-NADP-ME might partially dissociate into inactive homodimers and monomers. Additionally, under these conditions the C_4-NADP-ME is inhibited by malate. In sum, these two regulatory mechanisms would ensure reduced decarboxylase activity during the night, which is important to prevent CO_2 overproduction in conditions of reduced total photosynthetic activity.

Our findings provide evidence for the existence of an additional *in vivo* control mechanism of ZmC$_4$-NADP-ME activity, which is provided by phosphorylation of Ser419. Reversible phosphorylation of this residue provides fast and effective reduction of the ZmC$_4$-NADP-ME activity during the first hours of illumination. We hypothesize, that the physiological significance of this regulatory mechanism lies in reducing the CO_2 production under conditions when sufficient amounts of this molecule are provided from other sources (Figure 4).

First, respiration of CO_2 in maize BS cells was shown to aid the carbon concentrating mechanism under certain conditions (Bellasio and Griffiths, 2014b). Thus, a scenario is possible, in which during the first hours of illumination a contribution of this process to the CO_2 assimilation by RubisCO is significant enough, and a fully active ZmC$_4$-NADP-ME will only provoke a wasteful leakage of the CO_2 back to the M cells. In addition, in maize the PEP-CK was shown to function as a secondary decarboxylase in the BS cells (Pick et al., 2011), and the contribution to C$_4$ acids decarboxylation by these two enzymes was reported to be flexible, changing in response to different environmental conditions (Sharwood et al., 2014). In this regard, phosphorylation at Ser419 might lead to a rapid decrease of the C$_4$-NADP-ME activity, leading to an enhanced use of aspartate as a carbon carrier in the C$_4$ cycle in maize. This would accommodate the changes in the metabolic demands provoked by environmental changes.
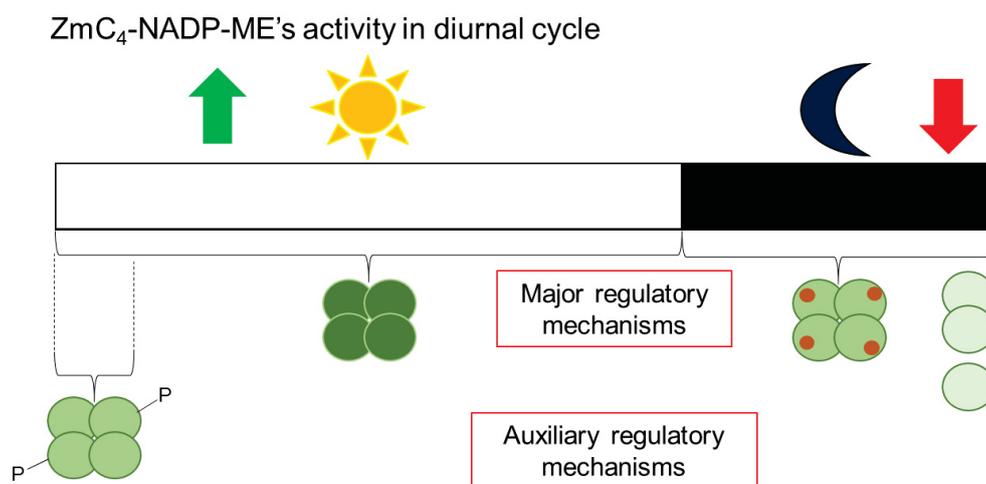
Figure 4: Illustration of the regulatory mechanisms affecting ZmC$_4$-NADP-ME's activity within diurnal cycle. Maximal activity of the ZmC$_4$-NADP-ME under illumination is enabled by its presence in a homotetrameric form. During the night, the enzymatic activity declines, which results from the decreased stability of the homoterameric assembly leading to formation of the inactive homodimers and monomers. In addition, enzymatic

activity would be further reduced by allosteric malate inhibition. During the first hours of illumination the reversible Ser419 phosphorylation would provide fast and effective reduction of the ZmC4-NADP-ME's activity.

7 References

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature **422**: 198–207

Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. Nature **537**: 347–355

Alvarez CE, Detarsio E, Moreno S, Andreo CS, Drincovich MF (2012) Functional characterization of residues involved in redox modulation of maize photosynthetic NADP-malic enzyme activity. Plant Cell Physiol **53**: 1144–1153

Alvarez CE, Saigo M, Margarit E, Andreo CS, Drincovich MF (2013) Kinetics and functional diversity among the five members of the NADP-malic enzyme family from *Zea mays*, a C$_4$ species. Photosynth Res **115**: 65–80

Anderson LE (1971) Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases. Biochim Biophys Acta **235**: 237–244

Andersson I (2008) Catalysis and regulation in Rubisco. J Exp Bot **59**: 1555–1568

Andreo CS, Gonzalez DH, Iglesias AA (1987) Higher plant phosphoenolpyruvate carboxylase. FEBS Lett **213**: 1–8

Arrivault S, Obata T, Szecówka M, Mengin V, Guenther M, Hoehne M, Fernie AR, Stitt M (2017) Metabolite pools and carbon flow during C$_4$ photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation. J Exp Bot **68**: 283–298

Artus NN, Edwards GE (1985) NAD-malic enzyme from plants. FEBS Lett **182**: 225–233

Ashton AR, Hatch MD (1983) Regulation of C$_4$ photosynthesis: regulation of pyruvate, P$_i$ dikinase by ADP-dependent phosphorylation and dephosphorylation. Biochem Biophys Res Comm **115**: 53–60

Aubry S, Brown NJ, Hibberd JM (2011) The role of proteins in C$_3$ plants prior to their recruitment into the C$_4$ pathway. J Exp Bot **62**: 3049–3059

Bakrim N, Echevarria C, Cretin C, Arrio-Dupont M, Pierre JN, Vidal J, Chollet R, Gadal P (1992) Regulatory phosphorylation of *Sorghum* leaf phospho*enol*pyruvate carboxylase: Identification of the protein-serine kinase and some elements of the signal-transduction cascade. Eur J Biochem **204**: 821–830

Barford D, Hu S-H, Johnson LN (1991) Structural mechanism for glycogen-phosphorylase control by phosphorylation and AMP. J Mol Biol **218**: 233–260

Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, et al (2017) UniProt: The universal protein knowledgebase. Nucleic Acids Res **45**: D158–D169

Bauwe H, Hagemann M, Fernie AR (2010) Photorespiration: players, partners and origin. Trends Plant Sci **15**: 330–336

Bellasio C, Griffiths H (2014a) The Operation of Two Decarboxylases, Transamination, and Partitioning of C$_4$ Metabolic Processes between Mesophyll and Bundle Sheath Cells Allows Light Capture To Be Balanced for the Maize C$_4$ Pathway. Plant Physiol **164**: 466–480

Bellasio C, Griffiths H (2014b) Acclimation to low light by C4 maize: Implications for bundle sheath leakiness. Plant, Cell Environ **37**: 1046–1058

Beltrao P, Bork P, Krogan NJ, Van Noort V (2013) Evolution and functional cross-talk of protein post-translational modifications. Mol Syst Biol **9**: 1–13

Bengoechea-Alonso MT, Ericsson J (2016) The phosphorylation-dependent regulation of nuclear SREBP1 during mitosis links lipid metabolism and cell growth. Cell Cycle **15**: 2753–2765

Bi YD, Wang HX, Lu TC, Li X hui, Shen Z, Chen YB, Wang BC (2011) Large-scale analysis of phosphorylated proteins in maize leaf. Planta **233**: 383–392

Bologna FP, Andreo CS, Drincovich MF (2007) *Escherichia coli* Malic Enzymes: Two Isoforms with Substantial Differences in Kinetic Properties, Metabolic Regulation, and Structure. J Bacteriol **189**: 5937–5946

Botha CEJ (1992) Plasmodesmatal distribution, structure and frequency in relation to assimilation in C$_3$ and C$_4$

grasses in southern Africa. Planta **187**: 348–358

Bräutigam A, Gowik U (2016) Photorespiration connects C$_3$ and C$_4$ photosynthesis. J Exp Bot **67**: 2953–2962

Bräutigam A, Hoffmann-Benning S, Weber APM (2008) Comparative Proteomics of Chloroplast Envelopes from C$_3$ and C$_4$ Plants Reveals Specific Adaptations of the Plastid Envelope to C$_4$ Photosynthesis and Candidate Proteins Required for Maintaining C$_4$ Metabolite Fluxe. Plant Physiol **148**: 568–579

Brodin P, Jojic V, Gao T, Bhattacharya S, Angel CJL, Furman D, Shen-Orr S, Dekker CL, Swan GE, Butte AJ, et al (2015) Variation in the human immune system is largely driven by non-heritable influences. Cell **160**: 37–47

Buchanan BB (1980) Role of Light in the Regulation of Chloroplast Enzymes. Annu Rev Plant Physiol **31**: 341–374

Buchanan BB, Balmer Y (2005) Redox Regulation: A Broadening Horizon. Annu Rev Plant Biol **56**: 187–220

Burnell JN, Chastain CJ (2006) Cloning and expression of maize-leaf pyruvate, P$_i$ dikinase regulatory protein gene. Biochem Biophys Res Commun **345**: 675–680

Burnell JN, Hatch MD (1986) Activation and Inactivation of an Enzyme Catalyzed by a Single, Bifunctional Protein: A New Example and Why. Arch Biochem Biophys **245**: 297–304

Burnell JN, Hatch MD (1985) Regulation of C$_4$ Photosynthesis : Purification and Properties of the Protein Catalyzing ADP-Mediated Inactivation and P$_i$-Mediated Activation of Pyruvate , P$_i$ Dikinase. **23**: 490–503

von Caemmerer S, Edwards GE, Koteyeva N, Cousins AB (2014) Single cell C$_4$ photosynthesis in aquatic and terrestrial plants: A gas exchange perspective. Aquat Bot **118**: 71–80

von Caemmerer S, Quick WP, Furbank RT (2012) The development of C$_4$ rice: Current progress and future challenges. Science (80-) **336**: 1671–1672

Cain JA, Solis N, Cordwell SJ (2014) Beyond gene expression: The impact of protein post-translational modifications in bacteria. J Proteomics **97**: 265–286

Chang GG, Tong L (2003) Structure and Function of Malic Enzymes, A New Class of Oxidative Decarboxylases. Biochemistry **42**: 12721–12733

Chang Y-M, Liu W-Y, Shih AC-C, Shen M-N, Lu C-H, Lu M-YJ, Yang H-W, Wang T-Y, Chen SC-C, Chen SM, et al (2012) Characterizing Regulatory and Functional Differentiation between Maize Mesophyll and Bundle Sheath Cells by Transcriptomic Analysis. Plant Physiol **160**: 165–177

Chao Q, Liu XY, Mei YC, Gao ZF, Chen YB, Qian CR, Hao YB, Wang BC (2014) Light-regulated phosphorylation of maize phosphoenolpyruvate carboxykinase plays a vital role in its activity. Plant Mol Biol **85**: 95–105

Chapman KSR, Hatch MD (1981) Aspartate decarboxylation in the bundle sheath cells of *Zea mays* and its possible contribution to C$_4$ photosynthesis. Aust J Plant Physiol **8**: 237–248

Chastain CJ, Xu W, Parsley K, Sarath G, Hibberd JM, Chollet R (2008) The pyruvate, orthophosphate dikinase regulatory proteins of Arabidopsis possess a novel, unprecedented Ser/Thr protein kinase primary structure. Plant J **53**: 854–863

Chen Y-B, Lu T-C, Wang H-X, Shen J, Bu T-T, Chao Q, Gao Z-F, Zhu X-G, Wang Y-F, Wang B-C (2014) Posttranslational Modification of Maize Chloroplast Pyruvate Orthophosphate Dikinase Reveals the Precise Regulatory Mechanism of Its Enzymatic Activity. Plant Physiol **165**: 534–549

Chen Z, Cole PA (2015) Synthetic approaches to protein phosphorylation. Curr Opin Chem Biol **28**: 115–122

Chi W, Yang J, Wu N, Zhang F (2004) Four rice genes encoding NADP malic enzyme exhibit distinct expression profiles. Biosci Biotechnol Biochem **68**: 1865–1874

Christin P-A, Osborne CP (2013) The recurrent assembly of C$_4$ photosynthesis, an evolutionary tale. Photosynth Res **117**: 163–175

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N (2008) Oligocene CO$_2$ Decline Promoted C$_4$ Photosynthesis in Grasses. Curr Biol **18**: 37–43

Christin PA, Samaritani E, Petitpierre B, Salamin N, Besnard G (2009) Evolutionary Insights on C$_4$ Photosynthetic Subtypes in Grasses from Genomics and Phylogenetics. Genome Biol Evol **1**: 221–230

Cieśla J, Fraczyk T, Rode W (2011) Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta Biochim Pol **58**: 137–148

Conceição P, Mendoza RU (2009) Anatomy of the global food crisis. Third World Q **30**: 1159–1182

Covshoff S, Burgess SJ, Kneřová J, Kümpers BMC (2014) Getting the most out of natural variation in C_4 photosynthesis. Photosynth Res **119**: 157–167

Covshoff S, Hibberd JM (2012) Integrating C_4 photosynthesis into C_3 crops to increase yield potential. Curr Opin Biotechnol **23**: 209–214

Danila FR, Quick WP, White RG, Furbank RT, von Caemmerer S (2016) The Metabolite Pathway between Bundle Sheath and Mesophyll: Quantification of Plasmodesmata in Leaves of C_3 and C_4 Monocots. Plant Cell **28**: 1461–1471

Day EK, Sosale NG, Lazzara MJ (2016) Cell signaling regulation by protein phosphorylation: a multivariate, heterogeneous, and context-dependent process. Curr Opin Biotechnol **40**: 185–192

Dephoure N, Gould KL, Gygi SP, Kellogg DR (2013) Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. Mol Biol Cell **24**: 535–542

Detarsio E, Alvarez CE, Saigo M, Andreo CS, Drincovich MF (2007) Identification of domains involved in tetramerization and malate inhibition of maize C_4-NADP-malic enzyme. J Biol Chem **282**: 6053–6060

Detarsio E, Maurino VG, Alvarez CE, Müller GL, Andreo CS, Drincovich MF (2008) Maize cytosolic NADP-malic enzyme (ZmCytNADP-ME): a phylogenetically distant isoform specifically expressed in embryo and emerging roots. Plant Mol Biol **68**: 355–367

Deutscher J, Saier MH (2005) Ser/Thr/Tyr Protein Phosphorylation in Bacteria - For Long Time Neglected, Now Well Established. J Mol Microbiol Biotechnol **9**: 125–131

Dittrich P, Campbell WH, Black CC (1973) Phosphoenolpyruvate Carboxykinase in Plants Exhibiting Crassulacean Acid Metabolism. Plant Physiol **52**: 357–361

Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. Science **312**: 212–217

Downton WJS, Hawker JS (1973) Enzymes of starch and sucrose metabolism in *Zea mays* leaves. Phytochemistry **12**: 1551–1556

Drincovich MF, Andreo CS (1994) Redox regulation of maize NADP-malic enzyme by thiol-disulfide interchange: effect of reduced thioredoxin on activity. Biochim Biophys Acta **1206**: 10–16

Drincovich MF, Casati P, Andreo CS (2001) NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. FEBS Lett **490**: 1–6

Drincovich MF, Iglesias A a, Andreo CS (1991) Interaction of divalent metal ions with the $NADP^+$-malic enzyme from maize leaves. Physiol Plant **81**: 462–466

Edwards EJ, Osborne CP, Stromberg CAE, Smith SA, Bond WJ, Christin PA, Cousins AB, Duvall MR, Fox DL, Freckleton RP, et al (2010) The Origins of C_4 Grasslands: Integrating Evolutionary and Ecosystem Science. Science **328**: 587–591

Edwards GE, Andreo CS (1992) NADP-Malic Enzyme From Plants. Phytochemistry **31**: 1845–1857

Edwards GE, Nakamoto H (1985) PYRUVATE, P_i DIKINASE AND NADP-MALATE DEHYDROGENASE IN C_4 PHOTOSYNTHESIS: Properties and Mechanism of Light/Dark Regulation. Annu Rev Plant Physiol **36**: 255–286

Erb TJ, Zarzycki J (2018) A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. Curr Opin Biotechnol **49**: 100–107

Fahnenstich H, Saigo M, Niessen M, Zanor MI, Andreo CS, Fernie AR, Drincovich MF, Flügge U-I, Maurino VG (2007) Alteration of organic acid metabolism in Arabidopsis overexpressing the maize C_4 NADP-malic enzyme causes accelerated senescence during extended darkness. Plant Physiol **145**: 640–652

Fesquet D, Labbé JC, Derancourt J, Capony JP, Galas S, Girard F, Lorca T, Shuttleworth J, Dorée M, Cavadore JC (1993) The MO15 gene encodes the catalytic subunit of a protein kinase that activates cdc2 and other cyclin-dependent kinases (CDKs) through phosphorylation of Thr161 and its homologues. EMBO J **12**: 3111–3121

Fischer EH, Krebs EG (1955) Conversion of phosphorylase b to phosphorylase a in muscle extracts. Proc Natl

Acad Sci **216**: 121–132

Fouracre JP, Ando S, Langdale JA (2014) Cracking the Kranz enigma with systems biology. J Exp Bot **65**: 3327–3339

Fry AM, O'Regan L, Sabir SR, Bayliss R (2012) Cell cycle regulation by the NEK family of protein kinases. J Cell Sci **125**: 4423–4433

Furbank RT (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? J Exp Bot **62**: 3103–3108

Furbank RT (2017) Walking the C_4 pathway: past, present, and future. J Exp Bot **68**: 4057–4066

Furbank RT, Stitt M, Foyer CH (1985) Intercellular compartmentation of sucrose synthesis in leaves of *Zea mays* L. Planta **164**: 172–178

Gerrard Wheeler MC, Arias CL, Tronconi MA, Maurino VG, Andreo CS, Drincovich MF (2008) *Arabidopsis thaliana* NADP-malic enzyme isoforms: high degree of identity but clearly distinct properties. Plant Mol Biol **67**: 231–242

Gerrard Wheeler MC, Drincovich FM, Andreo CS, Flugge U, Tronconi M a, Maurino VG (2005) A Comprehensive Analysis of the NADP-Malic Enzyme Gene Family of Arabidopsis. Plant Physiol **139**: 39–51

Gowik U, Westhoff P (2011) The Path from C_3 to C_4 Photosynthesis. Plant Physiol **155**: 56–63

Griffiths H, Weller G, Toy LFM, Dennis RJ (2013) You're so vein: bundle sheath physiology, phylogeny and evolution in C3 and C4 plants. Plant, Cell Environ **36**: 249–261

Gruber AV, Feiz L (2018) Rubisco Assembly in the Chloroplast. Front Mol Biosci **5**: 1–11

Gutierrez M, Gracen VE, Edwards GE (1974) Biochemical and cytological relationships in C_4 plants. Planta **119**: 279–300

Han X, Aslanian A, Yates JR (2008) Mass spectrometry for proteomics. Curr Opin Chem Biol **12**: 483–490

Harmel R, Fiedler D (2018) Features and regulation of non-enzymatic post-translational modifications. Nat Chem Biol **14**: 244–252

Hatch MD (1987) C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Arch Biochem Biophys **895**: 81–106

Hatch MD (1971) The C_4-pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C_4-dicarboxylic acid. Biochem J **125**: 425–432

Hatch MD, Burnell JN (1990) Carbonic Anhydrase Activity in Leaves and Its Role in the First Step of C_4 Photosynthesis. Plant Physiol **93**: 825–828

Hatch MD, Kagawa T, Craig S (1975) Subdivision of C_4-pathway Species Based on Differing C_4 Acid Decarboxylating Systems and Ultrastructural Features. Aust J Plant Physiol **2**: 111–128

Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. Biochem J **101**: 103–111

Heldt HW, Werdan K, Milovancev M, Geller G (1973) Alkalization of the chloroplast stroma caused by light-dependent proton flux into the thylakoid space. Biochim Biophys Acta **314**: 224–241

Hibberd JM, Covshoff S (2010) The Regulation of Gene Expression Required for C_4 Photosynthesis. Annu Rev Plant Biol **61**: 181–207

Humphrey SJ, James DE, Mann M (2015) Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. Trends Endocrinol Metab **26**: 676–687

Hunter T (2012) Why nature chose phosphate to modify proteins. Philos Trans R Soc B Biol Sci **367**: 2513–2516

Hurleys JH (1990) Regulation of isocitrate dehydrogenase by phosphorylation involves no long-range conformational change in the free enzyme. J Biol Chem **265**: 3599–3602

Iglesias AA, Andreo CS (1990) Kinetic and Structural Properties of NADP-Malic Enzyme from Sugarcane Leaves. Plant Physiol **92**: 66–72

Jensen ON (2006) Interpreting the protein language using proteomics. Nat Rev Mol Cell Biol **7**: 391–403

Jiang L, Chen Y-B, Zheng J, Chen Z, Liu Y, Tao Y, Wu W, Chen Z, Wang B-C (2016) Structural Basis of Reversible Phosphorylation by Maize Pyruvate Orthophosphate Dikinase Regulatory Protein. Plant Physiol **170**: 732–741

Jiao J, Chollet R (1988) Light/Dark Regulation of Maize Leaf Phosphoenolpyruvate Carboxylase by *in Vivo* Phosphorylation. Arch Biochem Biophys **261**: 409–417

Jiao J, Chollet R (1989) Regulatory Seryl-Phosphorylation of C$_4$ Phosphoenolpyruvate Carboxylase by a Soluble Protein Kinase from Maize Leaves. Arch Biochem Biophys **269**: 526–535

Jiao J, Chollet R (1990) Regulatory Phosphorylation of Serine-15 in Maize Phosphoenolpyruvate Carboxylase by a C$_4$-Leaf Protein-Serine Kinase. Arch Biochem Biophys **283**: 300–305

Jin J, Pawson T (2012) Modular evolution of phosphorylation-based signalling systems. Philos Trans R Soc B Biol Sci **367**: 2540–2555

Johnson LN (2009) The regulation of protein phosphorylation. Biochem Soc Trans **37**: 627–641

Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev **101**: 2209–2242

Jungblut PR, Holzhütter HG, Apweiler R, Schlüter H (2008) The speciation of the proteome. Chem Cent J **2**: 1–10

Jünger MA, Aebersold R (2014) Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. **3**: 83–112

Keeley JE, Rundel PW (2003) Evolution of CAM and C$_4$ Carbon-Concentrating Mechanisms. Int J Plant Sci **164**: S55–S77

Kromdijk J, Ubierna N, Cousins AB, Griffiths H (2014) Bundle-sheath leakiness in C$_4$ photosynthesis: a careful balancing act between CO$_2$ concentration and assimilation. J Exp Bot **65**: 3443–3457

Kusakina J, Dodd AN (2012) Phosphorylation in the plant circadian system. Trends Plant Sci **17**: 575–583

Langdale JA (2011) C$_4$ Cycles: Past, Present, and Future Research on C$_4$ Photosynthesis. Plant Cell **23**: 3879–3892

Langdale JA, Nelson T (1991) Spatial regulation of photosynthetic development in C4 plants. Trends Genet **7**: 191–196

Leakey ADB (2009) Rising atmospheric carbon dioxide concentration and the future of C$_4$ crops for food and fuel. Proc R Soc B Biol Sci **276**: 2333–2343

Lee H-J, Park Y-J, Seo PJ, Kim J-H, Sim H-J, Kim S-G, Park C-M (2015) Systemic Immunity Requires SnRK2.8-Mediated Nuclear Import of NPR1 in Arabidopsis. Plant Cell **27**: 3425–3438

Leegood RC (1985) The intercellular compartmentation of metabolites in leaves of *Zea mays* L. Planta **164**: 163–171

Leegood RC, Crowther D, Walker DA, Hind G (1983) Energetics of photosynthesis in *Zea mays*. I. Studies of the flash-induced electrochromic shift and fluorescence induction in bundle sheath cells. Biochim Biophys Acta **722**: 116–126

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al (2010) The developmental dynamics of the maize leaf transcriptome. Nat Genet **42**: 1060–1067

Li XS, Yuan BF, Feng YQ (2016) Recent advances in phosphopeptide enrichment: Strategies and techniques. Trends Anal Chem **78**: 70–83

Loeber G, Dworkin MB, Infante A, Ahorn H (1994a) Characterization of cytosolic malic enzyme in human tumor cells. FEBS Lett **344**: 181–186

Loeber G, Maurer-Fogy I, Schwendenwein R (1994b) Purification, cDNA cloning and heterologous expression of the human mitochondrial NADP$^+$-dependent malic enzyme. Biochem J **692**: 687–692

Long SP, Marshall-Colon A, Zhu XG (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell **161**: 56–66

Lothrop AP, Torres MP, Fuchs SM (2013) Deciphering post-translational modification codes. FEBS Lett **587**: 1247–1257

Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R (2018) Data-independent acquisition-

based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol **14**: e8126

Ludwig M (2013) Evolution of the C_4 photosynthetic pathway: events at the cellular and molecular levels. Photosynth Res **117**: 147–161

Ludwig M (2016) The Roles of Organic Acids in C_4 Photosynthesis. Front Plant Sci **7**: 1–11

Lunn JE, Furbank RT (1997) Localisation of sucrose-phosphate synthase and starch in leaves of C_4 plants. Planta **202**: 106–111

Lynch M, Conery JS (2000) The Evolutionary Fate and Consequences of Duplicate Genes. Science **290**: 1151–1155

Maier A, Zell MB, Maurino VG (2011) Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C_4 and C_3 photosynthesis. J Exp Bot **62**: 3061–3069

Majeran W, Cai Y, Sun Q, van Wijk KJ (2005) Functional Differentiation of Bundle Sheath and Mesophyll Maize Chloroplasts Determined by Comparative Proteomics. Plant Cell **17** 3111-3140

Majeran W, Zybailov B, Ytterberg a J, Dunsmore J, Sun Q, van Wijk KJ (2008) Consequences of C_4 Differentiation for Chloroplast Membrane Proteomes in Maize Mesophyll and Bundle Sheath Cells. Mol Cell Proteomics **7**: 1609–1638

Mandell DJ, Chorny I, Groban ES, Wong SE, Levine E, Rapp CS, Jacobson MP (2007) Strengths of hydrogen bonds involving phosphorylated amino acid side chains. J Am Chem Soc **129**: 820–827

Mann M, Ong S, Gr M, Steen H, Jensen ON, Pandey A (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol **20**: 261–268

Maurino VG, Drincovich MF, Casati P, Andreo CS, Edwards GE, Ku MSB, Gupta SK, Franceschi VR (1997) NADP-malic enzyme: Immunolocalization in different tissues of the C_4 plant maize and the C_3 plant wheat. J Exp Bot **48**: 799–811

Maurino VG, Saigo M, Andreo CS, Drincovich MF (2001) Non-photosynthetic "malic enzyme" from maize: a constituvely expressed enzyme that responds to plant defence inducers. Plant Mol Biol **45**: 409–420

Mazzucotelli E, Mastrangelo AM, Crosatti C, Guerra D, Stanca AM, Cattivelli L (2008) Abiotic stress response in plants: When post-transcriptional and post-translational regulations control transcription. Plant Sci **174**: 420–431

Meierhoff K, Westhoff P (1993) Differential biogenesis of photosystem II in mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants: the non-stoichiometric abundance of the subunits of photosystem II in the bundle-sheath chloroplasts and the translational activity of te plastome-encoded genes. Planta **191**: 23–33

Mertz RA, Brutnell TP (2014) Bundle sheath suberization in grass leaves: multiple barriers to characterization. J Exp Bot **65**: 3371–3380

Mijakovic I, Macek B (2012) Impact of phosphoproteomics on studies of bacterial physiology. FEMS Microbiol Rev **36**: 877–892

Monson RK (2003) Gene Duplication, Neofunctionalization, and the Evolution of C_4 Photosynthesis. **164**: S43-S54

Moorhead GBG, De Wever V, Templeton G, Kerk D (2009) Evolution of protein phosphatases in plants and animals. Biochem J **417**: 401–409

Moroney J V., Jungnick N, Dimario RJ, Longstreth DJ (2013) Photorespiration and carbon concentrating mechanisms: Two adaptations to high O_2, low CO_2 conditions. Photosynth Res **117**: 121–131

Munekage YN (2016) Light harvesting and chloroplast electron transport in NADP-malic enzyme type C_4 plants. Curr Opin Plant Biol **31**: 9–15

Nelson T (2011) The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C_4 leaves. J Exp Bot **62**: 3039–3048

Nigg E (1995) Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. Bioessays **17**: 471–480

Nimmo HG, Carter PA, Fewson CA, McNaughton GAL, Nimmo GA, Wilkins MB (1990) Regulation of phosphoenolpyruvate carboxylase: an example of signal transduction via protein phosphorylation in higher

plants. Adv Enzym Regul **30**: 121–126

Nishi H, Shaytan A, Panchenko AR (2014) Physicochemical mechanisms of protein regulation by phosphorylation. Front Genet **5**: 1–10

Nita-Lazar A, Saito-Benz H, White FM (2008) Quantitative phosphoproteomics by mass spectrometry: Past, present, and future. Proteomics **8**: 4433–4443

Nussinov R, Tsai CJ, Xin F, Radivojac P (2012) Allosteric post-translational modification codes. Trends Biochem Sci **37**: 447–455

Oliveira AP, Ludwig C, Picotti P, Kogadeeva M, Aebersold R, Sauer U (2012) Regulation of yeast central metabolism by enzyme phosphorylation. Mol Syst Biol **8**: 1–13

Olsen J V., Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. Cell **127**: 635–648

Paulus JK, Schlieper D, Groth G (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. Nat Commun **4**: 1517–1518

Pawson T, Gish GD, Nash P (2001) SH2 domains, interaction modules and cellular wiring. Trends Cell Biol **11**: 504–511

Peterhansel C, Maurino VG (2011) Photorespiration Redesigned. Plant Physiol **155**: 49–55

Pick TR, Brautigam A, Schluter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, et al (2011) Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C_4 Model and Provides Candidates for Regulation. Plant Cell **23**: 4208–4220

Raines CA (2003) The Calvin cycle revisited. Photosynth Res **75**: 1–10

Rao X, Dixon RA (2016) The Differences between NAD-ME and NADP-ME Subtypes of C_4 Photosynthesis: More than Decarboxylating Enzymes. Front Plant Sci **7**: 1–9

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS One **8**: 1–8

Reeves G, Grangé-Guermente MJ, Hibberd JM (2017) Regulatory gateways for cell-specific gene expression in C_4 leaves with Kranz anatomy. J Exp Bot **68**: 107–116

Richards AL, Merrill AE, Coon JJ (2015) Proteome sequencing goes deep. Curr Opin Chem Biol **24**: 11–17

Robles MS, Humphrey SJ, Mann M (2017) Phosphorylation Is a Central Mechanism for Circadian Control of Metabolism and Physiology. Cell Metab **25**: 118–127

Roskoski R (2015) A historical overview of protein kinases and their targeted small molecule inhibitors. Pharmacol Res **100**: 1–23

Sage RF (2016) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. J Exp Bot **67**: 4039–4056

Sage RF (2014) Stopping the leaks: New insights into C4 photosynthesis at low light. Plant, Cell Environ **37**: 1037–1041

Sage RF (2004) The evolution of C_4 photosynthesis. New Phytol **161**: 341–370

Sage RF, Khoshravesh R, Sage TL (2014) From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. J Exp Bot **65**: 3341–3356

Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the Evolution of C_4 Photosynthesis. Annu Rev Plant Biol **63**: 19–47

Sage RF, Stata M (2015) Photosynthetic diversity meets biodiversity: The C_4 plant example. J Plant Physiol **172**: 104–119

Sage RF, Zhu X-G (2011) Exploiting the engine of C_4 photosynthesis. J Exp Bot **62**: 2989–3000

Saigo M, Alvarez CE, Andreo CS, Drincovich MF (2013) Plastidial NADP-malic enzymes from grasses: Unraveling the way to the C_4 specific isoforms. Plant Physiol Biochem **63**: 39–48

Saigo M, Bologna F, Maurino VG, Detarsio E, Andreo CS, Drincovich MF (2004) Maize recombinant non-C_4 NADP-malic enzyme: A novel dimeric enzyme with high specific activity. Plant Mol Biol **55**: 97–107

Saleh A, Withers J, Mohan R, Marqués J, Gu Y, Yan S, Zavaliev R, Nomoto M, Tada Y, Dong X (2015) Posttranslational Modifications of the Master Transcriptional Regulator NPR1 Enable Dynamic but Tight Control of Plant Immune Responses. Cell Host Microbe **18**: 169–182

Salih E (2005) Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. Mass Spectrom Rev **24**: 828–846

Schlüter U, Denton AK, Bräutigam A (2016) Understanding metabolite transport and metabolism in C_4 plants through RNA-seq. Curr Opin Plant Biol **31**: 83–90

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Nature **326**: 1112–1115

Schulze WX, Usadel B (2010) Quantitation in Mass-Spectrometry-Based Proteomics. Annu Rev Plant Biol **61**: 491–516

Schwämmle V, Verano-Braga T, Roepstorff P (2015) Computational and statistical methods for high-throughput analysis of post-translational modifications of proteins. J Proteomics **129**: 3–15

Seet BT, Dikic I, Zhou MM, Pawson T (2006) Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol **7**: 473–483

Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski JR, Cox J, Mann M (2014) Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. Cell Rep **8**: 1583–1594

Sharpe RM, Offermann S (2014) One decade after the discovery of single-cell C_4 species in terrestrial plants: what did we learn about the minimal requirements of C_4 photosynthesis? Photosynth Res **119**: 169–180

Sharwood RE, Sonawane B V., Ghannoum O (2014) Photosynthetic flexibility in maize exposed to salinity and shade. J Exp Bot **65**: 3715–3724

Silva-Sanchez C, Li H, Chen S (2015) Recent advances and challenges in plant phosphoproteomics. Proteomics **15**: 1127–1141

Sims RJ, Reinberg D (2008) Is there a code embedded in proteins that is based on post-translational modifications? Nat Rev Mol Cell Biol **9**: 815–820

Smoly I, Shemesh N, Ziv-Ukelson M, Ben-Zvi A, Yeger-Lotem E (2017) An Asymmetrically Balanced Organization of Kinases versus Phosphatases across Eukaryotes Determines Their Distinct Impacts. PLoS Comput Biol **13**: 1–21

Spoel SH (2018) Orchestrating the proteome with post-translational modifications. J Exp Bot **69**: 4499–4503

Spoel SH, Mou Z, Tada Y, Spivey NW, Genschik P, Dong X (2009) Proteasome-Mediated Turnover of the Transcription Coactivator NPR1 Plays Dual Roles in Regulating Plant Immunity. Cell **137**: 860–872

Stata M, Sage TL, Rennie TD, Khoshravesh R, Sultmanis S, Khaikin Y, Ludwig M, Sage RF (2014) Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants. Plant Cell Environ **37**: 2587–2600

Steen H, Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol **5**: 699–711

Still CJ, Berry JA, Collatz GJ, DeFries RS (2003) Global distribution of C_3 and C_4 vegetation: Carbon cycle implications. Global Biogeochem Cycles **17**: 6-1-6–14

Stitt M, Heldt HW (1985) Generation and maintenance of concentration gradients between the mesophyll and bundle sheath in maize leaves. BBA - Bioenerg **808**: 400–414

Stitt M, Zhu XG (2014) The large pools of metabolites involved in intercellular metabolite shuttles in C_4 photosynthesis provide enormous flexibility and robustness in a fluctuating light environment. Plant, Cell Environ **37**: 1985–1988

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. Mol Syst Biol **4**: 1–7

Sun F, Ding Y, Ji Q, Liang Z, Deng X, Wong CCL, Yi C, Zhang L, Xie S, Alvarez S, et al (2012) Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. Proc Natl Acad Sci **109**: 15461–15466

Tausta LS, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. J Exp Bot **65**: 3543–3555

Tausta SL, Coyle HM, Rothermel B, Stiefel V, Nelson T (2002) Maize C4 and non-C4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. Plant Mol Biol **50**: 635–652

Thorsness PE, Koshland DE (1987) Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. J Biol Chem **262**: 10422–10425

Turkan I, Uzilday B, Dietz KJ, Bräutigam A, Ozgur R (2018) Reactive oxygen species and redox regulation in mesophyll and bundle sheath cells of C4 plants. J Exp Bot **69**: 3321–3331

Ueno Y, Imanari E, Emura J, Yoshizawa-kumagaye K, Nakajima K, Inami K (2000) Immunological analysis of the phosphorylation state of maize C4-form phosphoenolpyruvate carboxylase with specific antibodies raised against a synthetic phosphorylated peptide. Plant J **21**: 17–26

Venne AS, Kollipara L, Zahedi RP (2014) The next level of complexity: Crosstalk of posttranslational modifications. Proteomics **14**: 513–524

Vorapreeda T, Thammarongtham C, Cheevadhanarak S, Laoteng K (2013) Repertoire of malic enzymes in yeast and fungi: insight into their evolutionary functional and structural significance. Microbiology **159**: 2548–2557

Walker RP, Acheson RM, Técsi LI, Leegood RC (1997) Phospho*enol*pyruvate Carboxykinase in C4 Plants: Its Role and Regulation. Aust J Agric Res **24**: 459–468

Walker RP, Chen Z, Acheson RM, Leegood RC (2002) Effects of Phosphorylation on Phospho*enol*pyruvate Carboxykinase from the C4 Plant Guinea Grass. Plant Physiol **128**: 165–172

Walker RP, Leegood RC (1996) Phosphorylation of phosphoenolpyruvate carboxykinase in plants. Biochem J **317**: 653–658

Walsh CT, Garneau-Tsodikova S, Gatto GJ (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. Angew Chemie - Int Ed **44**: 7342–7372

Wang L, Peterson RB, Brutnell TP (2011) Regulatory mechanisms underlying C4 photosynthesis. New Phytol **190**: 9–20

Wang S, Li Z, Shen H, Zhang Z, Yin Y, Wang Q, Zhao X, Ji J (2016) Quantitative Phosphoproteomic Study Reveals that Protein Kinase A Regulates Neural Stem Cell Differentiation Through Phosphorylation of Catenin Beta-1 and Glycogen Synthase Kinase 3β. Stem Cells **34**: 2090–2101

Wang Y, Bräutigam A, Weber APM, Zhu XG (2014) Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. J Exp Bot **65**: 3567–3578

Weber APM, Bräutigam A (2013) The role of membrane transport in metabolic engineering of plant primary metabolism. Curr Opin Biotechnol **24**: 256–262

Weber APM, von Caemmerer S (2010) Plastid transport and metabolism of C3 and C4 plants - comparative analysis and possible biotechnological exploitation. Curr Opin Plant Biol **13**: 257–265

Wedding RT (1989) Malic Enzymes of Higher Plants. Plant Physiol **90**: 367–371

Werdan K, Heldt HW, Milovancev M (1975) The role of pH in the regulation of carbon fixation in the chloroplast stroma. Studies on CO2 fixation in the light and dark. BBA - Bioenerg **396**: 276–292

Westhoff P, Gowik U (2004) Evolution of C4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. Ann Bot **93**: 13–23

Wheeler T, von Braun J (2013) Climate change impacts on global food security. Science **341**: 508–513

Whitney SM, Houtz RL, Alonso H (2011) Advancing Our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco. Plant Physiol **155**: 27–35

Williams BP, Aubry S, Hibberd JM (2012) Molecular evolution of genes recruited into C4 photosynthesis. Trends Plant Sci **17**: 213–220

Winter K, Foster JG, Edwards GE, Holtum JA (1982) Intracellular Localization of Enzymes of Carbon Metabolism in *Mesembryanthemum crystallinum* Exhibiting C3 Photosynthetic Characteristics or

Performing Crassulacean Acid Metabolism. Plant Physiol **69**: 300–307

Witze ESS, Old WMM, Resing KAA, Ahn NGG (2007) Mapping protein post-translational modifications with mass spectrometry. Nat Methods **4**: 798–806

Woo KC, Anderson JM, Boardman NK, Downton WJ, Osmond CB, Thorne SW (1970) Deficient Photosystem II in Agranal Bundle Sheath Chloroplasts of C$_4$ Plants. Proc Natl Acad Sci USA **67**: 18–25

Xin F, Radivojac P (2012) Post-translational modifications induce significant yet not extreme changes to protein structure. Bioinformatics **28**: 2905–2913

Xu J, Zhang S (2015) Mitogen-activated protein kinase cascades in signaling plant growth and development. Trends Plant Sci **20**: 56–64

Zhu L, Li N (2013) Quantitation, networking, and function of protein phosphorylation in plant cell. Front Plant Sci **3**: 1–7

Zhu XG, Long SP, Ort DR (2008) What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Curr Opin Biotechnol **19**: 153–159

Zhulin IB, Nikolskaya AN, Galperin MY (2003) Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in Bacteria and Archaea. J Bacteriol **185**: 285–294

Zolnierowicz S, Bollen M (2000) Protein phosphorylation and protein phosphatases. EMBO J **19**: 483–488

Acknowledgements

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst, und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt habe, und dass alle Ausführungen, die anderen Werken wörtlich oder sinngemäß entnommen wurden, als solche kenntlich gemacht und im Literaturverzeichnis vollständig hinterlegt worden sind.

Diese Dissertation habe ich in dieser oder ähnlicher Form noch bei keiner anderen Fakultät vorgelegt.

Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 29.01.2019 _____

Anastasiia Bovdilova