# Molecular characterization of control systems for small molecule damage in photosynthetic eukaryotes

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

**Meike Hüdig**

aus Velbert

Düsseldorf, Oktober 2018

aus dem Institut für Entwicklungs- und Molekularbiologie der Pflanzen

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. PD Dr. Veronica G. Maurino

2. Prof. Dr. Martin J. Lercher

Tag der mündlichen Prüfung:

11.12.2018

# Contents

## Eidestattliche Versicherung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Meike Hüdig

## Abbreviations

| | |
|---|---|
| 2-KG | 2-ketoglutarate |
| AA | amino acid |
| BSC | bundle sheath cell |
| CETCH | crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA |
| CoA | coenzyme A |
| COG3236 | DUF1768 containing protein |
| DHAP | dihydroxyacetone phosphate |
| DTT | dithiothreitol |
| DUF | domain of unknown function |
| EC | enzyme commission number |
| G3P | glyceraldehyde-3-phosphate |
| GAPDH | glyceraldehyde-3-phosphate dehydrogenase |
| GLX | glyoxalase |
| GlyDH | glycolate dehydrogenase |
| GOX | glycolate oxidase |
| GSH | glutathione |
| GST | glutathione (*S*)-transferase |
| GTP | guanosine triphosphate |
| HPR | hydroxypyruvate reductase |
| MCL | β-methylmalyl-CoA lyase |
| MCT | malyl-CoA thioester hydrolase |
| Met-*R*-O | methionine-(*R*)-sulfoxide |
| mMDH | mitochondrial malate dehydrogenase |
| MOPS | 3-(*N*-morpholino)propanesulfonic acid |
| MSA | multiple sequence alignment |
| MSRB | methionine sulfoxide reductase B |
| mya | million years ago |
| NAD | nicotinamide adenine dinucleotide |
| NAD(P)HX | NAD(P)H-hydrate |
| NAD-ME | NAD-dependent malic enzyme |
| NADP | nicotinamide adenine dinucleotide phosphate |
| NADP-ME | NADP-dependent malic enzyme |
| OAA | oxaloacetate |
| PEP | phospho*enol*pyruvate |
| PEPC | phospho*enol*pyruvate carboxylase |
| PEPCK | phospho*enol*pyruvate carboxykinase |
| PG | 2-phosphoglycolate |
| PGA | 3-phosphoglycerate |
| PMSF | phenylmethylsulfonyl fluoride |
| PPDK | pyruvate phosphate dikinase |
| qRT-PCR | quantitative real-time polymerase chain reaction |
| RCS | reactive carbonyl species |
| ROS | reactive oxygen species |
| RubisCO | ribulose-1,5-bisphosphate carboxylase/oxygenase |
| SDS-PAGE | sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| TPI | triose-phosphate isomerase |
| WGD | whole genome duplications |

## Summary

Cell metabolism in photosynthetic eukaryotes constitutes a complex network that involves both spontaneous chemical and enzymatically catalyzed reactions. This network of biochemical and chemical reactions is far from perfect. As a consequence, small molecules like cofactors, coenzymes, inorganic molecules, and metabolic intermediates are constantly exposed to two major sources of damage during metabolism - enzymatic errors and chemical damage.

A molecule that was damaged can be useless or toxic to the cell in a given condition. Harmful molecules damage other metabolites, macromolecules or inhibit important enzymes. If damaged molecules represent unusable dead-end metabolites, they drain energy and carbon equivalents from the metabolism. Thus, damage control systems that constantly repair or prevent damage in the cell have evolved. We described three categories of small molecule damage control in photosynthetic eukaryotes. Firstly, repair mechanisms convert damaged molecules back to ones that can be used in normal metabolism. Secondly, scavenging systems act on highly reactive molecules like RCS and ROS to convert them quickly and locally to less harmless molecules. Thirdly, metabolite steering systems prevent the formation of a damaged molecule by actively changing availability of substrate or co-factor pools.

So far, this research has been focused on bacteria, yeast, and mammalian systems, as they are potentially easier to study or attract more attention. Therefore, this thesis aimed to address a combination of the overlooked aspects of damage research: photosynthetic eukaryotes and small molecule damage control. Finding new aspects of known damage control systems and the discovery of new ones was approached. All three aspects of damage control, repair, scavenging, and steering, were studied in different photosynthetic eukaryotes. These damage control systems for plants were also reviewed (Hüdig *et al.*, 2018).

Damage repair systems were addressed in two cases: firstly, the discovery and description of the single step repair mechanism for L-2-hydroxyglutarate (L-2HG) by L-2-hydroxyglutarate dehydrogenase (L-2HGDH; EC 1.1.99.2) and its integration in the mitochondrial metabolism in *Arabidopsis thaliana* (Hüdig *et al.*, 2015; manuscript 2.1). Secondly, the characterization of an alternative glycolate oxidation pathway as part of the multi-step repair system photorespiration in the diatom *Phaeodactylum tricornutum* was discovered (Schmitz *et al.*, 2017*a*; manuscript 2.2).

Damage control through scavenging systems was studies for the small RCS methylglyoxal. The molecular characterization of all enzymatic components for the local scavenging within the plant cell was described in detail for the model plant *Arabidopsis thaliana* (Schmitz *et al.*, 2017*b*; manuscript 2.3). New aspects of the multiform steering system $C_4$ photosynthesis were studied to discover the mechanisms behind the recruitment of NAD-dependent malic enzyme (NAD-ME) for $C_4$ biochemistry. The molecular evolution and regulation of the NAD-ME $C_4$ subtype pathway in the genus *Cleome* was described in manuscript draft 2.5 (Hüdig *et al.*, in preparation).

Taken together these manuscripts advanced the field of small molecule damage control by discovering new damage control systems in photosynthetic organisms, describing these pathways in detail, and expanding the knowledge on known mechanisms by offering new perspectives and finding new molecular characteristics.

# 1. Introduction

Cell metabolism in photosynthetic eukaryotes constitutes a complex network. It consists of biochemical reactions that are compartmentalized, but not detached from one another. Usually, when considering cell metabolism, one means the entire, well-defined and studied, enzyme-based network (Michal, 2014). However, it is evident that this cell metabolism does not only consist of enzyme-catalyzed reactions. Metabolism once began in a cell-like structure without the presence of proteins, especially enzymes. The chemical reactions that occurred did not vanish, if they can take place in ambient conditions, they will (Lerma-Ortiz *et al.*, 2016). These spontaneous reactions extent the metabolic network of the cell and thus strongly influences the cellular environment (Keller *et al.*, 2015).

## 1.1 Molecule damage through metabolic activity

Metabolism involves both spontaneous and enzymatically catalyzed reactions, which might overlap in the cell, meaning the same reaction can occur through both mechanisms (Keller *et al.*, 2015). This network of biochemical and chemical reactions is far from perfect. As a consequence, small molecules like cofactors, coenzymes, inorganic molecules, and metabolic intermediates are constantly exposed to two major sources of damage during metabolism - enzymatic errors and chemical damage (Hanson *et al.*, 2016). A molecule that was damaged can be useless or toxic to the cell in a given condition. Harmful molecules damage other metabolites, macromolecules or inhibit important enzymes. If damaged molecules represent unusable dead-end metabolites, they drain energy and carbon equivalents from the metabolism (Linster *et al.*, 2013).

This illustrates that the status of a certain molecule depends on its impact on cellular metabolism. It is not per se damaged, but the context defines it as such. A metabolite occurring in one cell type or one cellular compartment might not be harmful while in others it is wasteful or toxic. Likewise, a certain concentration can be the threshold for a damaged molecule. Therefore, controlling the formation of damaged molecules is a crucial part of the cellular metabolism, a view that has gained additional attention in the last years (Sun *et al.*, 2017).

## 1.2 The complexity of damaged molecules

As described above, the production and effects of a damaged molecule strongly depends on the cellular metabolic status and a given condition. In the following chapters, examples of damage reactions and their consequences on cellular metabolism are presented. An overview on molecule damage is outlined in figure 1.

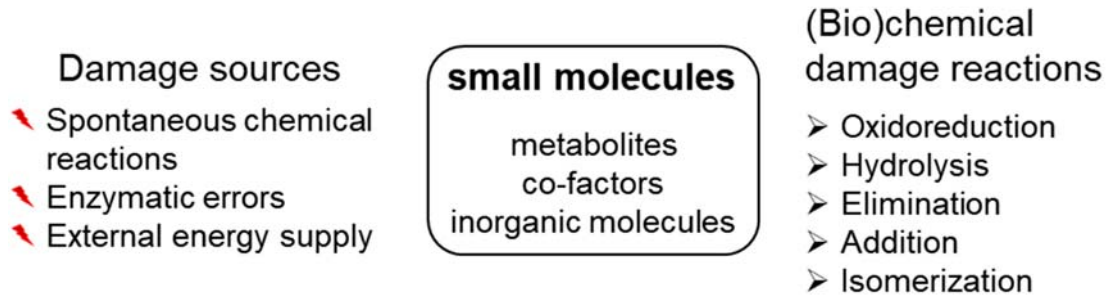



**Figure 1.** Small molecules are exposed to different sources of damage which results in a multitude of (bio)chemical reactions. These turn undamaged molecules into damaged ones with varying degrees of severity.

### 1.2.1. Molecule damage through spontaneous reactions

To categorize spontaneous reactions and also their enzyme-based counterparts, the classification according to the enzyme commission number (EC) is used: oxidoreduction (EC 1), transfer (EC 2), hydrolysis (EC 3), decomposition (EC 4), isomerization (EC 5), ligation (EC 6), and translocation (EC 7; updated September 2018) (Bairoch, 2000). Not every spontaneous reaction leads to the formation of a damaged molecule. Nevertheless, un-controlled reactions are more prone to damage molecules through their inherent spontaneous nature. Examples of spontaneous reactions occurring during cellular activity are those leading to the formation of active toxic species, such as reactive oxygen species (ROS) and reactive carbonyl species (RCS) (Apel & Hirt, 2004; Mano, 2012).

ROS arise constantly during normal metabolism and are produced in excess under certain stress conditions. They comprise oxygen-derived molecules such as hydrogen peroxide ($H_2O_2$), the superoxide anion, or hydroxyl radical, which are able to react quickly with other metabolites and macromolecules in the cell (Piedrafita *et al.*, 2015). $H_2O_2$ is the most stable and diffusible form of ROS. Its spontaneous degradation via free $Fe^{2+}$-mediated Fenton reaction into superoxide and hydroxyl radical occurs in aqueous solutions and can cause severe damage to the cell (Stadtman, 1993; Tachiev *et al.*, 2000).

RCS are a group of small molecules described as α,β-unsaturated carbonyls. This unsaturated carbon bond makes RCS extremely electrophilic and therefore highly reactive to a wide range of molecules (Farmer & Davoine, 2007; Mano, 2012). A special case, in which a highly toxic RCS forms, is the spontaneous isomerization of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (G3P) during cellular activity leading to the formation of methylglyoxal (Schmitz *et al.*, 2017*b*). Formation of methylglyoxal can also occur by action of triose-phosphate isomerase (TPI, EC 5.3.1.1), when the phosphate group is spontaneously eliminated from the enediol phosphate intermediate (Richard, 1993). Various other RCS are released through spontaneous decomposition of lipid peroxide radicals. This links RCS and ROS in a dynamic manner, as ROS are usually the cause of oxidative (stress) conditions in the cell (Mano, 2012). Hence, RCS cause severe damage to cellular macromolecules and therefore they impair metabolism (Vistoli *et al.*, 2013).

Spontaneous hydration of NADH and NADPH to NAD(P)HX at cellular pH, which is accelerated at higher temperatures, causes a major drain from the cofactor pools (Oppenheimer & Kaplan, 1974; Yoshida & Dave, 1975; Marbaix *et al.*, 2011). NAD(P)HX strongly inhibit dehydrogenases of the oxidative pentose phosphate pathway and glycerol-3-phosphate dehydrogenase and therefore must be eliminated (Yoshida & Dave, 1975; Prabhakar *et al.*, 1998).

Several other small molecules can spontaneously react with metabolites and macromolecules in the cell. These reactions create random adducts that need to be removed, often at the expense of special repair enzymes and energy (Lerma-Ortiz *et al.*, 2016). Glutathione (GSH) can be transferred actively to other metabolites (e.g. glutathione *S*-transferases, EC 2.5.1.18), but its thiol group also reacts spontaneously at cellular pH. These reactions are enhanced with increasing concentrations of GSH, leading to increasing concentrations of GSH adducts (Satoh, 1995).

### 1.2.2. Molecule damage through enzymatic reactions

Metabolites can also be damaged by the activity of promiscuous enzymes that act on multiple substrates or catalyze abnormal reactions on their regular substrate. Enzyme promiscuity is needed and often found in secondary metabolism (e.g. cytochrome P450s), where substrates variation is large and the investment in highly specified enzymes is inefficient (Khersonsky & S.Tawfik, 2010). Substrate switches can also be enabled by changes in co-factors such as metal ions (Schmitz *et al.*, 2017*b*). Cellular

availability of these co-factors will thus influence the reaction outcome and consequently the formation of damaged molecules.

One infamous example and thereby illustration of enzymes that catalyze side reactions on their main substrate is Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). It either carboxylates (EC 4.1.1.39) or oxygenates (EC 1.13.11) ribulose-1,5-bisphosphate yielding in the latter case the toxic 2-phosphoglycolate (PG) next to PGA. Similar to the availability of inorganic and organic co-factors, cellular concentration of carbon dioxide and oxygen influences the rate of metabolite damage caused by RubisCO (Hanson, 2016).

### 1.2.3. Molecule damage is increased by external factors

In addition to the intracellular determinants of metabolite damage, e.g. concentration of reaction partners or enzyme promiscuity, external factors influence the formation of damaged metabolites. All of them have in common that in one way or another they introduce excess energy into a relatively closed system, the cell. Increased temperature affects the biochemical system broadly and leads to faster reactions. This reduces the activation energy of prior thermodynamically inhibited spontaneous reactions and leads to an increase in variety and rate of damaged metabolites due to increase in reactions rates and decline in specificity of enzymes. A narrower effect is produced by any type of radiation inflicted on the cell. Ultra violet radiation is a constant and fluctuating damage source not only to metabolites, but naturally to all cell components (Landry *et al.*, 1995). Ionizing radiation damages all cellular components and is a source for ROS generation leading to molecule damage (Stadtman, 1993).

Other external stress factors include abiotic and biotic factors that influence cellular status through induction of one or several of the mentioned damage routes. Abiotic stresses such as osmotic stresses, salt, and heavy metal stress, induce unfavorable conditions for homeostasis or ROS dependent pathways thus leading to the formation of damaged molecules (Apel & Hirt, 2004). Uptake of ions from the environment leads also to adducts on macromolecules and an increase of ion mediated spontaneous reactions (Cuypers *et al.*, 2004). Biotic factors like pathogens or herbivores usually cause signaling cascades within the cells that lead to damage up to cell death. In addition, cells produce actively damaging conditions, e.g. $H_2O_2$ release, to cause damage to pathogenic intruders (Torres, 2010). Again, illustrating that even when the

underlying mechanism of damage to molecules is the same, context provides the status of a toxic or harmful molecule.

## 1.3 Molecule damage control systems

When molecule damage is inevitable, so is the existence of damage control systems. An emerging view on metabolite repair and damage pre-emption as part of the central cellular metabolism opens a wide field of studies (Linster *et al.*, 2013; Van Schaftingen *et al.*, 2013). Contrasting the two damage sources and the plethora of factors that influence the severity of the metabolite damage, three areas of molecule damage prevention are defined (**Figure 2**):

Firstly, repair mechanisms convert damaged molecules back to ones that can be used in normal metabolism. Secondly, scavenging systems act on highly reactive molecules like RCS and ROS to convert them quickly and locally to less harmless molecules. Thirdly, metabolite steering systems prevent the formation of a damaged molecule by actively changing availability of substrate or co-factor pools.

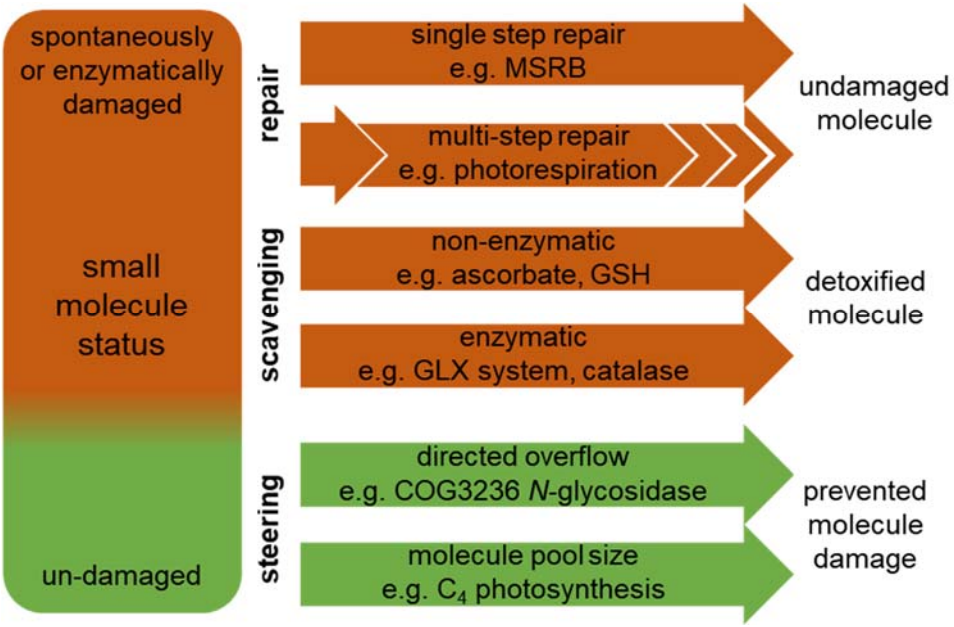



**Figure 2.** Molecule damage control systems range from elimination (repair) to prevention mechanisms (metabolite steering). Repair systems return a damaged small molecule to a usable metabolite. Scavenging systems convert reactive species produced during metabolism into more harmless ones and/or control their concentration to enable additional function in signaling. Steering mechanisms by-pass the production of damaged small molecules by controlling pool size of critical substrates (reduction or enrichment). MSRB, methionine sulfoxide reductase B; GSH, glutathione; GLX, glyoxalase; COG3236, DUF1768 containing protein involved in riboflavin biosynthesis. Modified from Hüdig *et al.* (2018).

### 1.3.1 Repair mechanisms

Enzymatic repair systems convert dead-end and harmful metabolites back to metabolites used in the normal metabolic network. These repair systems exist in a wide range of forms. They can consist of simple one-step reactions or scale up to complex multi-step reactions involving multiple compartments of the cell (Bauwe *et al.*, 2010; Hanson *et al.*, 2016). Single step repair mechanisms either convert the damaged molecule back to the original one or convert it to an energetically favorable molecule and broaden thereby the metabolic network (see below). The latter concept has brought recent attention to the field of repair mechanism discovery, as these additional links in metabolism can be the starting point for synthetically designed metabolic pathways (Schwander *et al.*, 2016; Sun *et al.*, 2017). Multi-step repair mechanisms are inherently more complex and can offer multiple branching points. Hence, the process is completed once the original undamaged molecule is returned.

#### *Single step repair mechanisms*

An example for a single step repair mechanism that reverses the continuous occurring damage to the original molecule is the reduction of methionine-(*R*)-sulfoxide (Met-*R*-O) to methionine. Under oxidizing conditions, e.g. elevated ROS levels, free and protein bound methionine can be damaged spontaneously by metal ion catalyzed oxidation forming methionine sulfoxide (Stadtman, 1993). For both stereoisomers independent repair enzymes have been described. In plants methionine sulfoxide reductase B (MSRB; EC 1.8.4.12) uses thioredoxin to produce methionine from Met-*R*-O. Plant cells, as well as yeast and bacteria, lacking this repair mechanism have been shown to be less resistant to oxidative stress conditions (Vieira Dos Santos *et al.*, 2005; Le *et al.*, 2013).

Another option to repair a damaged molecule is to directly convert it into other harmless and required products instead of returning it to the original state. Malyl-CoA thioester hydrolase (MCT, EC 3.1.2.30) acts as a repair enzyme by hydrolysing malyl-CoA to malate and CoA, feeding these products directly into normal metabolism. Malyl-CoA is formed by a promiscuous reaction of the enzyme β-methylmalyl-CoA lyase (MCL) in the photosynthetic *Rhodobacter sphaeroides* (Erb *et al.*, 2010). In the synthetic *in vitro* CETCH cycle of $CO_2$ fixation, MCT is identically used to directly convert the side product of MCL to the desired end product of the cycle: malate (Schwander *et al.*, 2016; Sun *et al.*, 2017).

### *Multi-step repair mechanisms*

A two-step repair mechanism in photosynthetic organisms is the repair of the above-mentioned spontaneously formed cofactor hydrates NAD(P)HX. Additionally, glyceraldehyde-3-phosphate dehydrogenase (GAPDH, EC 1.2.1.12) as part of glycolysis produces NADHX in a side reaction (Rafter et al, 1954). GAPDH-mediated or spontaneous hydration of NAD(P)H yields natural racemates of (*R,S*)-NAD(P)HX. It has been shown that an epimerase converts (*R*)-NAD(P)HX to (*S*)-NAD(P)HX which in turn is converted back to NAD(P)H by the NAD(P)HX dehydratase (EC 4.2.1.93) (Marbaix *et al.*, 2011). This pathway is present in both chloroplast and mitochondria in higher plants (Colinas *et al.*, 2014).

The oxidative photosynthetic carbon cycle, or photorespiratory pathway, is a multi-step repair mechanism to deal with the highly toxic product of the oxygenase reaction of RubisCO. The generated damaged molecule PG is converted back to PGA in a series of enzymatic steps that involve multiple compartments of the cell (Maurino & Peterhänsel, 2010). The exact enzymatic steps and their subcellular location vary from unicellular organisms like the cyanobacterium Synechocystis, single celled eukaryotes like green alga Chlamydomonas, or diatoms, to higher plants, thus illustrating evolutionary flexibility and differing metabolic needs when dealing with the repair of PG (Nakamura *et al.*, 2005; Eisenhut *et al.*, 2008; Peterhänsel & Maurino, 2011; Schmitz *et al.*, 2017*a*).

The plant photorespiratory pathway includes chloroplasts, mitochondria, peroxisomes and the cytosol with at least 11 enzymes and additional transporters for the intermediates of the cycle. Additionally, there are alternative pathways for some of the reactions, e.g. the reduction of hydroxypyruvate to glycerate by hydroxypyruvate reductase (HPR), which takes place both in the peroxisome and cytosol (Timm *et al.*, 2008). In green algae of the chlorophyta linage, glycolate to glyoxylate conversion takes place in mitochondria and is performed by glycolate dehydrogenase (GlyDH, EC 1.1.99.14) (Nakamura *et al.*, 2005; Esser *et al.*, 2014). In green algae of the charophyte linage and higher plants glycolate oxidation takes place in the peroxisomes by glycolate oxidase (GOX, EC 1.1.3.15) (Esser *et al.*, 2014). Cyanobacteria like Synechocystis, albeit having a carbon concentration mechanism that reduces PG generation, inherit an active PG detoxification pathway and perform glycolate conversion to glyoxylate in the cytosol using GlyDH (Eisenhut *et al.*, 2008).

### 1.3.2 Scavenging systems

ROS and RCS are formed during normal metabolism and their production increase when plant cells are challenged by external conditions. ROS in plant cells are mainly produced in chloroplasts and mitochondria because of their high rate of electron flow, and also in peroxisomes by various metabolic pathways and their oxidases, e.g. glycolate oxidase as part of the photorespiratory pathway (Asada, 2006; Murphy, 2009; Sandalio *et al.*, 2013). Additional sources are the cytochrome P450s in cytoplasm and the endoplasmic reticulum, pH-dependent peroxidases at the cell wall and different oxidases located in the apoplast (Gill & Tuteja, 2010). When ROS are not immediately scavenged, radical species, especially the hydroxyl radical, can attack lipids. This ubiquitous lipid peroxidation throughout the cell serves as a source of RCS. Additionally, continuous RCS production from the TPI reaction in cytosol and chloroplasts is present in all organisms performing oxygenic photosynthesis (Gill & Tuteja, 2010; Schmitz *et al.*, 2017*b*).

The basal detoxification of ROS and RCS is achieved through non-enzymatic scavenging. Plants have large pools of ascorbate (up to 20 mM), tocopherols and GSH that exceed the concentration of produced highly reactive species under steady state conditions. This enables the cell to detoxify these highly reactive molecules in a quick and local manner. Highly reactive singlet oxygen is scavenged by multiple molecules namely ascorbate, carotenoids and tocopherol. Neither of the scavenging molecules is exclusive to only one form of ROS, they are also linked by their decreasing redox potential. Regeneration of the scavenger molecule is dependent on both pool size and ambient conditions, e.g. subjected to changes in cellular pH. Maintaining these pools of reduced scavenging molecules in proximity to the endogenous production sites is often included in the term ‘redox homeostasis’ of the cell (Foyer & Noctor, 2005).

Enzymatic scavenging deals either directly with detoxification of ROS and RCS or with the regeneration of the primary non-enzymatic scavenging molecules in the ROS and RCS pathways. Examples of enzymatic detoxification of ROS are superoxide dismutase (EC 1.15.1.1) acting on superoxide and catalase, ascorbate peroxidase (EC 1.11.1.11) and glutathione peroxidase (EC 1.11.1.9) acting on $H_2O_2$ (Gill & Tuteja, 2010). Glutathione (S)-transferases (GSTs, EC 2.5.1.18) act on a variety of electrophiles, including RCS from lipid peroxidation (Vistoli *et al.*, 2013).

The majority of the scavenging enzymes act on the quick and constant regeneration of ascorbate (water-water cycle), GSH, and regeneration of reduction equivalents necessary in these cycles (Asada, 2006; Gill & Tuteja, 2010). For example, monodehydroascorbate reductase (EC 1.6.5.4), dehydroascorbate reductase (EC 1.8.5.1) and glutathione reductase (EC 1.8.1.7) conjointly regenerate ascorbate and GSH involved in ROS detoxification. The GSH that scavenges RCS, including methylglyoxal and glyoxal, is regenerated by the GLX system involving (*S*)-D-lactoylglutathione lyase or glyoxalase I (GLXI; EC 4.4.1.5) and (*S*)-2-hydroxyacylglutathione hydrolase or glyoxalase II (GLXII; EC 3.1.2.6) (Thornalley, 1990). These enzymatic cycles add an additional layer of damage mitigation to the basal detoxification. Also, they enable the cell to react to increased levels of ROS and RCS production upon stress through upregulation, providing additional damage control when needed.

### 1.3.3 Steering systems

The third type of damage control systems, steering systems, are conceptually different from the ones described above. It is defined to comprise those systems that evolved to avoid the formation of damaged molecules. Steering systems can be classified by two mechanisms: directed overflow and molecule pool size control. In the first case, molecules that when available in excess, are prone to damage, need to be metabolized. They are diverted out of their canonical metabolic pathways and converted to less harmful ones. In the second case, the accumulation of molecules that can be used as substrates and co-factors of promiscuous enzymes is controlled to avoid their damage (Hüdig *et al.*, 2018).

#### *Directed overflow*

Only a few examples labeled as direct overflow mechanisms that are described in detail are to be found in literature. They all have in common that transcriptional, posttranslational or product feedback regulation for enzymes that produce highly reactive molecules is absent and does not prevent the accumulation of these damage-prone products (Reaves *et al.*, 2013). These intermediates are unstable and will spontaneously be converted to toxic products or the products themselves will end up being damaged.

During riboflavin biosynthesis in both the green lineage and in bacteria, un-controlled built-up of the first two highly unstable intermediates occurs and a steering enzyme has been identified. The first enzyme GTP cyclohydrolase II (EC 3.5.4.25) causes the continuous production of the first intermediate due to a lack of a feedback mechanism. The accumulation of the intermediates leads to an increase in damaged molecules due to the formation of Maillard products from the intermediates or direct yield of reactive nucleophiles due to spontaneous decomposition. COG3236 proteins with the domain of unknown function (DUF) 1768 are *N*-glycosidases that hydrolyze the reactive intermediates to ribose-5-phosphate and a pyrimidine, more harmless metabolites (Frelin *et al.*, 2015).

### *Molecule pool size control*

To prevent the production of damaged molecules, steering of molecule pools within the cell can be used. One repeated cause of damage to plant cells is the oxygen in the ambient air. As described before, RubisCOs oxygenase reaction yields PG which needs to be detoxified in the elaborate multi-step repair pathway termed photorespiration. Another option to prevent the accumulation of PG is to suppress the oxygenase reaction by enhancing the carboxylase reaction of RubisCO. Several independent mechanisms have evolved in photosynthetic organisms to locally concentrate $CO_2$ around RubisCO, one of these is $C_4$ photosynthesis (Hatch & Slack, 1966; Sage *et al.*, 2012).

$C_4$ photosynthesis is an umbrella term for a specific set of biochemical, anatomical and regulatory changes that result in the concentration of $CO_2$ around RubisCO up to 10 times higher than in cells without $CO_2$ enrichment. In $C_4$ photosynthetic plants, this concentration is achieved by spatial separation of pre-fixation of $CO_2$ and its release at the site of RubisCO. The pre-fixation is performed by a non-promiscuous enzyme phospho*enol*pyruvate carboxylase (PEPC, EC 4.1.1.31) that produces oxaloacetate (OAA) from phospho*enol*pyruvate (PEP) and $HCO_3^-$. OAA is than converted and transported to the RubisCO containing cell or cellular compartment. The release of $CO_2$ is managed by different decarboxylases depending on the variant of $C_4$ photosynthesis present in the species. Mainly NADP-malic enzyme (NADP-ME; EC 1.1.1.40), mitochondrial NAD-malic enzyme (NAD-ME; EC 1.1.1.39), and phospho*enol*pyruvate carboxykinase (PEPCK; EC 4.1.1.32) contribute, usually in concert, to the carboxylation reaction on either malate by the malic enzymes or OAA

by PEPCK (Furbank, 2011; von Caemmerer & Furbank, 2016). The resulting pyruvate is than shuttled back, in some variants involving conversion to and shuttling of alanine, to the site of PEP regeneration by pyruvate phosphate dikinase (PPDK; EC 2.7.9.1) (Wang *et al.*, 2014*b*). This elaborate cycle for $CO_2$ concentration evolved at least 61 times independently, spanning 19 plant families with approx. 8100 species performing a version of $C_4$ photosynthesis (Sage, 2017). The actual rewiring of the metabolism involves not only changes to the anatomy, biochemistry, and special regulation of involved proteins, but is also costly in terms of energy demand (Hatch, 1988; Wang *et al.*, 2014*a*). Nevertheless, plants in habitats where the oxygenase reaction of RubisCO would be enhanced, have repeatedly evolved this steering mechanism to minimize the damaging reaction. The evolutionary steps on the way were shown to involve steering of many metabolite pools, e.g. glycine and its decarboxylation, which all per se result in a better plant fitness, essentially leveling the pathway for this mechanism (Heckmann *et al.*, 2013; Bräutigam & Gowik, 2016).

## 1.4 Aim of the thesis

Cellular damage is a broad term that comprises different core areas, perspectives and methodological approaches. Damage to cellular macromolecules, such as nucleic acids, proteins, and lipids has been extensively investigated (Warner *et al.*, 1987; Gardner, 1989; Grossman *et al.*, 1998; Mary *et al.*, 2004). So far, this research has been focused on bacteria, yeast, and mammalian systems, as they are potentially easier to study or attract more attention.

Therefore, this thesis aims to address a combination of the overlooked aspects of damage research: photosynthetic eukaryotes and small molecule damage control. Finding new aspects of known damage control systems and the discovery of new ones will be approached. All three aspects of damage control, repair, scavenging, and steering, will be studied in different photosynthetic eukaryotes. These damage control systems for plants will be reviewed in manuscript 2.4 (Hüdig *et al.*, 2018).

Damage repair systems will be addressed in two cases: firstly, the discovery and description of the single step repair mechanism for L-2-hydroxyglutarate (L-2HG) by L-2-hydroxyglutarate dehydrogenase (L-2HGDH; EC 1.1.99.2) and its integration in the mitochondrial metabolism in *Arabidopsis thaliana* (Hüdig *et al.*, 2015; manuscript 2.1). Secondly, the characterization of an alternative glycolate oxidation pathway as

part of the multi-step repair system photorespiration in the diatom *Phaeodactylum tricornutum* will be described (Schmitz *et al.*, 2017*a*; manuscript 2.2).

Damage control through scavenging systems will also be addressed for the small RCS methylglyoxal. The molecular characterization of all enzymatic components for the local scavenging within the plant cell will be described (Schmitz *et al.*, 2017*b*; manuscript 2.3).

New aspects of the multiform steering system $C_4$ photosynthesis will be studied to discover the mechanisms behind the recruitment of NAD-ME for $C_4$ biochemistry. The molecular evolution and regulation of the NAD-ME $C_4$ subtype pathway in the genus *Cleome* will be described in manuscript draft 2.5 (Hüdig *et al.*, in preparation).

## 2. Manuscripts

### 2.1. Plants possess a cyclic mitochondrial metabolic pathway similar to the mammalian metabolic repair mechanism involving malate dehydrogenase and L-2-hydroxyglutarate dehydrogenase

Title: Plants possess a cyclic mitochondrial metabolic pathway similar to the mammalian metabolic repair mechanism involving malate dehydrogenase and L-2-hydroxyglutarate dehydrogenase

Authors: Meike Hüdig*, Alexander Maier*, Isabell Scherrers, Laura Seidel, Erwin E.W. Jansen, Tabea Mettler-Altmann, Martin K.M. Engqvist and Veronica G. Maurino

Published in Plant Cell Physiol. 56(9):1820–1830 on 21.07.2015

* these authors contributed equally to the work

Contributions

- Performed experiments: Heterologous protein expression, kinetic measurements, tissue expression analysis
- Development of methodology: assay for new substrates of mitochondrial malate dehydrogenase
- Data analysis and interpretation
- Figure production
- Wrote part of and edited the final manuscript

License Number 4451840692229 (for print only)

Full reference:

Hüdig M, Maier A, Scherrers I, Seidel L, Jansen EEW, Mettler-Altmann T, Engqvist MKM and Maurino VG. (2015) Plants possess a cyclic mitochondrial metabolic pathway similar to the mammalian metabolic repair mechanism Involving malate dehydrogenase and L-2-hydroxyglutarate dehydrogenase. *Plant & Cell Physiology* **56**, 1820–1830.

Permanent link via DOI:

https://doi.org/10.1093/pcp/pcv108

## 2.2. The ancestors of diatoms evolved a unique mitochondrial dehydrogenase to oxidize photorespiratory glycolate

Title: The ancestors of diatoms evolved a unique mitochondrial dehydrogenase to oxidize photorespiratory glycolate

Authors: Jessica Schmitz*, Nishtala V. Srikanth*, Meike Hüdig, Gereon Poschmann, Martin J. Lercher, Veronica G. Maurino

* these authors contributed equally to the work

Published in Photosynth. Res. 132:183–196 on 28.02.2017

Contributions:

- Performed subcellular localization
- Development of methodology: microscopy
- Analysis of raw data and interpretation
- Figure editing
- Edited manuscript

License Number 4451840967057 (Print and electronic license)

Full reference:

Schmitz J, Srikanth NV, Hüdig M, Poschmann G, Lercher MJ and Maurino VG. (2017) The ancestors of diatoms evolved a unique mitochondrial dehydrogenase to oxidize photorespiratory glycolate. *Photosynthesis Research* **132**, 183–196.

Permanent link via DOI:

https://doi.org/10.1007/s11120-017-0355-1

## 2.3. Defense against reactive carbonyl species involves at least three subcellular compartments where individual components of the system respond to cellular sugar status

Title: Defense against reactive carbonyl species involves at least three subcellular compartments where individual components of the system respond to cellular sugar status

Authors: Jessica Schmitz, Isabell Charis Dittmar, Jörn Dennis Brockmann, Marc Schmidt, Meike Hüdig, Alessandro W. Rossoni and Veronica G. Maurino

Published in Plant Cell 29:3234–3254 originally published online on 17.11.2017

Contributions:

- Raw data acquisition, validation, replication of experiments and analysis of subcellular localization
- Figure editing
- Wrote parts of and edited full manuscript

License Id: 4451850467592

Full reference:

Schmitz J, Dittmar IC, Brockmann JD, Schmidt M, Hüdig M, Rossoni AW and Maurino VG. (2017) Defense against reactive carbonyl species involves at least three subcellular compartments where individual components of the system respond to cellular sugar status. *Plant Cell* **29**, 3234–3254.

Permanent link via DOI:

https://doi.org/10.1105/tpc.17.00258

## 2.4. Biochemical control systems for small molecule damage in plants

Title: Biochemical control systems for small molecule damage in plants

Authors: Meike Hüdig, Jessica Schmitz, M. K. M. Engqvist and Veronica G. Maurino

Published in Plant Signal. Behav. 13:5, e1477906 on 26.06.2018

Contributions:

- Evolution of ideas
- Figure draft
- Wrote parts of and edited full manuscript

License:



Full reference:

Hüdig M, Schmitz J, Engqvist MKM and Maurino VG. (2018) Biochemical control systems for small molecule damage in plants. *Plant Signaling & Behavior* **13**, 1–7.

Permanent link via DOI:

https://doi.org/10.1080/15592324.2018.1477906

## 2.5. Evolution of NAD-ME from a TCA cycle-associated enzyme to a $C_4$ decarboxylase was aided by duplication of a β-subunit in the genus Cleome

Manuscript draft

Title: Evolution of NAD-ME from a TCA cycle-associated enzyme to a $C_4$ decarboxylase was aided by duplication of a β-subunit in the genus Cleome

Authors: Meike Hüdig, Marcos A. Tronconi, Gereon Poschmann and Veronica G. Maurino

Contributions:

- Performing of experiments: NAD-ME sequence acquisition, phylogenetic analysis, cloning, heterologous protein production and characterization, kinetic analysis, plant growth and harvest, native and SDS-PAGE analysis, in gel malic enzyme activity assay, qRT-PCR, structural 3D modelling
- Processing and analysis of raw data
- Interpretation of data
- Evolution of hypothesis and research goals
- Figure production
- Wrote part of and edited full manuscript

# Evolution of NAD-malic enzyme from a TCA cycle-associated enzyme to a $C_4$ decarboxylase was aided by duplication of a β-subunit in the genus Cleome

Meike Hüdig[1], Marcos A. Tronconi[2], Gereon Poschmann[3], Veronica G. Maurino[1]

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

[2]Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET). Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, 2000 Rosario, Argentina.

[3]Molecular Proteomics Laboratory, Center for Biological and Medical Research (BMFZ), Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany.

Running title: Molecular Evolution of $C_4$-NAD-malic enzyme

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, Universitätsstraße 1, 40225, Düsseldorf, Germany. Tel.: 49-211-8112368. Fax: 49-211-8113706.

## Abstract

In plants, NAD-malic enzyme (NAD-ME; EC 1.1.1.39) is exclusively present in mitochondria where it is involved in malate respiration. In species performing a specific type of $C_4$ photosynthesis, NAD-ME was co-opted to decarboxylate malate in bundle sheath cell mitochondria to facilitate $CO_2$ concentration in close proximity to RubisCO. A specific $C_4$ NAD-ME was never identified to date. Here, we use Cleomaceae as model system to study the molecular evolution of NAD-ME in the emergence of $C_4$ photosynthesis. We found that all three genes coding for NAD-ME proteins in the $C_4$ Cleome *Gynandropsis gynandra* are expressed at higher levels than their $C_3$ orthologs in *Tarenaya hassleriana*, thus transcriptional analysis did not allow the identification of a $C_4$ candidate. Biochemical and proteomic analyses showed that a unique NAD-ME entity exists exclusively in photosynthetic tissues of *G. gynandra*. This heterodimeric NAD-ME is composed of *Gg*NAD-ME1 and *Gg*NAD-ME2 and exhibits high affinity towards malate and high catalytic activity. In addition to this heterodimer, an NAD-ME entity composed of all three subunits exists in *G. gynandra* photosynthetic and heterotrophic tissues. We also identified an NAD-ME entity composed of all three subunits in photosynthetic and heterotrophic tissues of *T. hassleriana* with similar kinetic properties as the enzyme found in the corresponding tissues in $C_4$ Cleome. We propose that these NAD-MEs composed of three subunits perform the housekeeping functions in $C_3$ and $C_4$ Cleome tissues. Finally, we show that *Gg*NAD-ME2, one of two retained β NAD-ME subunits, is inactive and likely evolved to mediate the $C_4$ function. Our results lead to the conclusion that in the $C_4$ Cleome species the functions of NAD-ME as a basic metabolic enzyme and as a specific photosynthetic decarboxylase are performed by separated enzymatic entities originated through a differential combination of subunits.

## Introduction

Many of the world's most productive crops are species bearing the $C_4$ photosynthetic metabolism as an adaptation to high light intensities, temperatures and drought. $C_4$ plants have evolved biochemical pumps to concentrate $CO_2$ at the site of Rubisco, which decrease the oxygenation reaction and, thus, limit the wasteful flux through photorespiration (Furbank & Hatch, 1987). Compared with plants using the ancestral $C_3$ photosynthetic pathway, $C_4$ plants show higher water and nitrogen use efficiency, which allows increased productivity in warm habitats (Edwards *et al.*, 2010).

The transition from $C_3$ to $C_4$ photosynthesis involved complex alterations to leaf anatomy and biochemistry. This occurred in at least 62 linages of angiosperms independently and thus is an example of convergent evolution (Sage *et al.*, 2011). In $C_4$ plants, $CO_2$ concentration is achieved by a two-step mechanism. First, phospho*enol*pyruvate carboxylase (PEPC) produces oxaloacetate (OAA) in the cytosol (primary $CO_2$ fixation as $C_4$ acid). Then, OAA is converted to malate and/or aspartate, and transported to the site of Rubisco, where $CO_2$ is released again and the final $CO_2$ fixation takes place (Christopher & Holtum, 1996; Drincovich *et al.*, 2011). Most $C_4$ plants developed a spatial separation of the biochemical components of the $CO_2$ pump by utilizing two adjacent cell types. In these plants, PEPC is located in the cytosol of mesophyll cells, and the formed $C_4$ acid is shuttled to the bundle sheath cells (BSCs) (Drincovich *et al.*, 2011). The release of $CO_2$ from malate in BSCs is mediated mainly by two different decarboxylases: NADP-malic enzyme (NADP-ME, EC 1.1.1.40) and NAD-ME (EC 1.1.1.39). $C_4$ plants are traditionally grouped into biochemical subtypes depending on the major decarboxylase used. About 40 $C_4$ lineages contain species using NADP-ME as their primary decarboxylase, while NAD-ME is used by species from about 20 lineages (Sage *et al.*, 2011). In either of these subtypes there exist the facultative addition of phospho*enol*pyruvate carboxykinase (PEPCK), which activity also releases $CO_2$ in BSCs (Furbank, 2011; Wang *et al.*, 2014*a*).

The $C_4$ decarboxylases did not evolve *de novo*. Instead, they were recruited from existing housekeeping isoforms (Monson, 2003; Aubry *et al.*, 2011; Maier *et al.*, 2011; Christin *et al.*, 2013). The existence of gene duplications is widely believed to be a pre-condition for the evolution of $C_4$ metabolism, as gene duplicates allow the plant to retain one gene copy for the ancestral (here: housekeeping) functions, while its duplicate is relieved from negative selection and suitable for acquiring substitutions advantageous for the evolving $C_4$ function (Monson, 2003). A specific $C_4$ NADP-ME, located in BSC chloroplasts, exists in all investigated species of the NADP-ME subtype (Matsuoka *et al.*, 2001; Drincovich *et al.*, 2011; Lauterbach *et al.*, 2017). This $C_4$ isoform originated by gene duplication of an existing chloroplastic non-photosynthetic NADP-ME and the subsequent neofunctionalization of one gene copy (Tausta *et al.*, 2002; Saigo *et al.*, 2004). Interestingly, to date there are no reports on the identification and characterization of a specific $C_4$ NAD-ME.

In plants, NAD-MEs are exclusively present in mitochondria where they are involved in malate respiration (Grover *et al.*, 1981; Tronconi *et al.*, 2008). In their housekeeping

roles, NAD-MEs are homo- or heterodimers of two distinct but phylogenetically related subunits, which shear 65% identity at the amino acid level: α-subunits with molecular masses of ~65 kDa and β-subunits with molecular masses of ~58 kDa (Maier *et al.*, 2011; Tronconi *et al.*, 2018). In *A. thaliana*, *AtNAD-ME1* (At4g13560) encodes the α-subunit, while *AtNAD-ME2* (At4g00570) encodes the β-subunit (Tronconi *et al.*, 2008). The separated recombinant homodimers and the reconstituted heterodimer present similar catalytic efficiencies but differ in kinetic mechanisms and their regulation by metabolic effectors (Tronconi *et al.*, 2008, 2010*a*, 2010*b*, 2015). Although *in vivo* it functions mostly as a heterodimer, the *At*NAD-ME2 homodimer is the only form of the enzyme present in anthers of Arabidopsis where its function would be coupled to glycolysis (Tronconi *et al.*, 2010*b*). The different biochemical properties of the NAD-ME enzymatic entities produced by alternative association of the subunits suggest that NAD-ME activity may be regulated by variations in the native association *in vivo* (Tronconi *et al.*, 2010*b*).

Biochemical analyses in some $C_4$ species indicated that NAD-ME from the dicotyledoneous *Amaranthus tricolor* and the grasses *Eleusine coracana* and *Panicum dichotomiflorum* are octamers of only one type of subunit (Ohsugi & Murata, 1980; Murata *et al.*, 1989), while those of *Amaranthus hypochondriacus* and *Amaranthus edulis* are composed of α- and β-subunits (Long *et al.*, 1994; Dever *et al.*, 1998). These enzymes were always regarded as the major $C_4$ photosynthetic isoforms, and it is intriguing that there exist no references to housekeeping isoforms of NAD-ME in $C_4$ plants. Intriguingly, comparative transcriptomes of leaves of the closely related Cleome *Tarenaya hassleriana*, a $C_3$ ornamental species, and *Gynandropsis gynandra*, an NAD-ME $C_4$ vegetable crop species, did not reveal any $C_4$-specific NAD-ME transcript (Bräutigam *et al.*, 2011). There is no clear evidence for the existence of a specific $C_4$ NAD-ME from data available. Therefore, it has been hypothesized that in $C_4$ plants of the NAD-ME subtype there may exist only one NAD-ME isoform, which may carry out both the housekeeping role in all cell types and the photosynthetic role exclusively in BSCs, where it is most abundantly expressed and probably subjected to metabolic or post transcriptional regulation (Maier *et al.*, 2011).

Cleomaceae is an excellent model system to study the molecular evolution of NAD-ME in the emergence of $C_4$ photosynthesis (Bayat *et al.*, 2018). Cleomaceae are closely related to Brassicaceae (Hall *et al.*, 2002), which diverged from each other around 24.2–49.4 million years ago (mya) (Schranz & Mitchell-Olds, 2006). Although

Brassicaceae (e.g., *A. thaliana*) and most Cleomaceae (e.g., *T. hassleriana)* are $C_3$ plants, NAD-ME subtype $C_4$ photosynthesis evolved *de novo* in Cleomaceae at least three times independently in *G. gynandra*, *Cleome oxalidea*, and *Cleome angustifolia* (Feodorova *et al.,* 2010*;* Koteyeva *et al.,* 2011). Interestingly, the ancestor of the $C_3$ *T. hassleriana* and $C_4$ *G. gynandra* Cleome species underwent a whole genome duplication approximately 20.0 mya (Schranz & Mitchell-Olds, 2006; Barker *et al.*, 2009). Specifically, a gene duplication of the ancestral NAD-ME β-subunit is retained in both *Cleome* species.

In this work, we address the question if a unique NAD-ME entity performs a dual function in $C_4$ plants or if indeed a $C_4$ specific isoform exist. By combining gene expression pattern and proteomic analyses and biochemical protein characterization, we found that in the $C_4$ Cleome species the functions of NAD-ME as a basic metabolic enzyme and as a specific photosynthetic decarboxylase are performed by separated enzymatic entities originated through a differential combination of subunits. Our data indicates that the duplication of the NAD-ME β-subunit was a potentiating evolutionary event that aided the repeated evolution of NAD-ME as a $C_4$ decarboxylase in this genus.

## Results

### Evolutionary history of NAD-ME proteins

To perform a comprehensive phylogenetic analysis of the mitochondrial NAD-ME in the green lineage, we gathered a total of 173 protein sequences from 70 species. These span from green algae, including the picoplancton *Micromonas pusilla* (Worden *et al.*, 2009) and *Chara braunii* as a recently sequenced representative of the Charophyta (Nishiyama *et al.*, 2018), classic representatives of the Spermatophytes' sister groups, mosses, and lycophytes, to 160 of the sequences from 59 angiosperm species.

We found that all investigated green algae species of the chlorophyte linage possess only one copy of *NAD-ME*, but gene duplications in the green algae of the charophyte lineage *C. braunii* and in the moss *Physcomitrella patens* have occurred (Fig. 1A). Further along the evolutionary trajectory, a gene duplication of *NAD-ME* has occurred between the split of the Lycophyta from the angiosperms which has been retained in all flowering plant species investigated. This makes the resulting paralogous NAD-ME

proteins in the angiosperms monophyletic (Fig. 1A, green and orange boxes). Whether this split into the so called α NAD-ME and β NAD-ME subunits occurred before or after the split of gymnosperms and angiosperms remains to be determined, as so far no NAD-ME sequences from gymnosperms are available.

We found that the 59 angiosperm species analyzed contain at least one α-subunit and one β-subunit each (Suppl. Fig. 1). Additional gene duplications have occurred in several angiosperms, resulting in 71 α NAD-ME and 89 β NAD-ME sequences, respectively. This demonstrates that if gene duplications or WGD occur, the gene coding for a β NAD-ME is more likely to be retained.

The focus of this study was the NAD-ME of the $C_4$ model species *Gynandropsis gynandra* from the genus Cleome. One of three independent $C_4$ origins that arose in the Cleomaceae, a sister family to the Brassicaceae that include the model species *A. thaliana*. We analyzed the NAD-ME sequences of several species of the Brassicales, including species of the Brassicaceae and Cleomaceae family in a separate phylogenetic analysis (Fig. 1B). Strikingly, the sequences of the $C_4$ Cleome *G. gynandra* show longer branches on the phylogenetic tree compared to the evolutionary close $C_3$ relative *T. hassleriana*. Additionally, we found *Gg*NAD-ME2 confidently (100% bootstrap support) placed as a sister group to all other β NAD-ME of the Brassicales (Fig. 1B, orange box). This most probably does not reflect the correct evolutionary relationship, which would place all Cleome β NAD-ME as a monophyletic group (cf. relationship of α NAD-ME), but is most probably caused by an unusual high amino acid exchange rate in the *Gg*NAD-ME2 sequence (see also results of the *in silico* analysis).

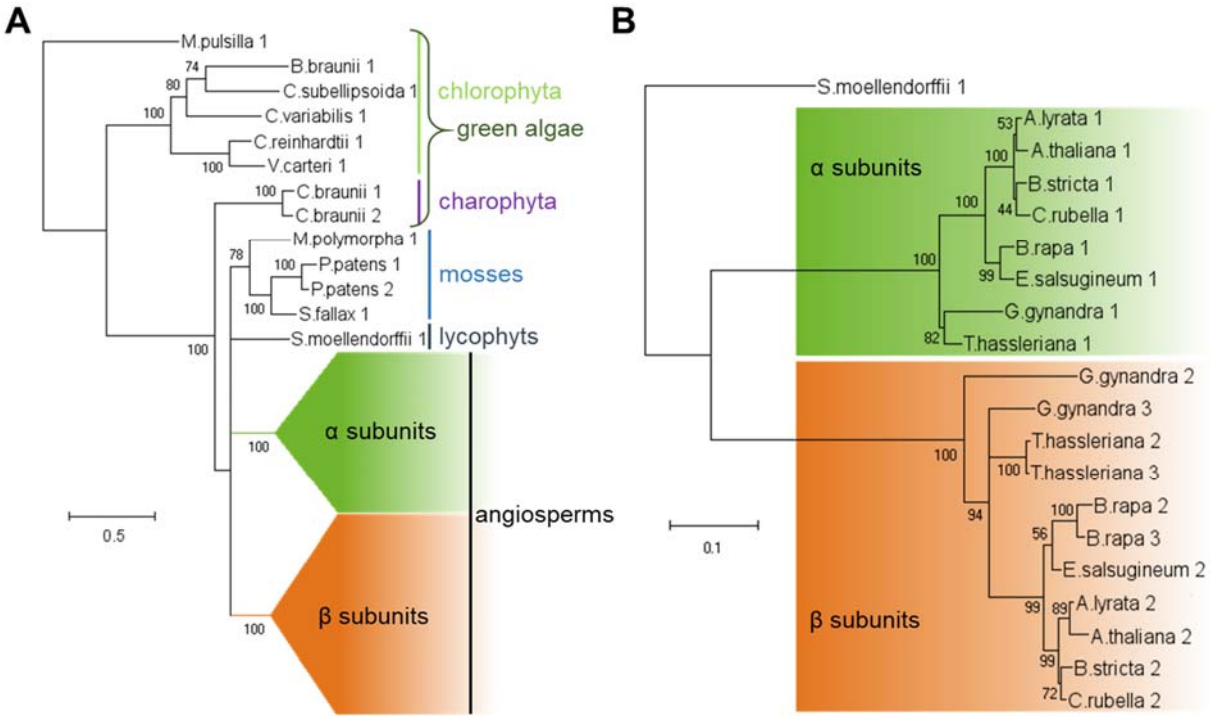



**Figure 1.** Maximum likelihood trees of NAD-ME sequences in the green lineage. Evolutionary history was inferred using the JTT+frequencies model. Numbers on branches represent bootstrap (n=1000) support values. Scale bar represents amino acid substitutions per site. A. Simplified representation of the phylogenetic tree of 173 sequences of the green lineage (full version in Suppl. Fig. 1). Rate differences were modeled with gamma parameter (G = 0.8060) and invariant sites (I = 7.60% sites). Branches with low support (<60%) were collapsed. B. Selected subtree of Brassicales sequences. The tree is rooted by an outgroup and all branches are shown with bootstrap support values. Rate differences were modeled with parameters specific for this subtree (G = 0.8733, I = 23.11% sites).

### Differential expression of *NAD-ME* genes in leaves of $C_3$ and $C_4$ Cleome species

A straightforward rationale is to assume that if a gene was recruited into the $C_4$ pathway it should be expressed at higher levels than in $C_3$ species and its transcript mostly accumulated during the day in photosynthetic organs (Christin *et al.*, 2013; Moreno-Villena *et al.*, 2018). To find such comparatively high expressed candidate gene(s) for $C_4$ NAD-ME, we analyzed the diel expression pattern of all *NAD-ME* genes in fully expanded leaves of *T. hassleriana* and *G. gynandra* (Fig. 2).

Quantitative transcriptional analysis of *NAD-ME* genes in leaves of $C_3$ *T. hassleriana* revealed that *ThNAD-ME1* is significantly higher expressed than *ThNAD-ME2* (28-fold higher, 8h) and *ThNAD-ME3* (17-fold higher, 8h) at all sampled time points during the diel course ($p<0.001$). *ThNAD-ME1* expression slightly increases 1.6-fold from early

morning (1 h) to midday (8 h) ($p<0.01$; Fig. 2A) while *ThNAD-ME2* and *ThNAD-ME3* expression levels do not change along the diel course ($p>0.05$).

Analysis of expression of *NAD-ME* genes in leaves of $C_4$ *G. gynandra* indicated similar expression levels ($p>0.05$) of all three genes during the night to day transition period (23h to 1h; Fig. 2B). The expression levels of *GgNAD-ME2* and *GgNAD-ME3* are maintained at those similar values during the whole day-night course (Fig. 2B). *GgNAD-ME1* expression exceeds the expression of *GgNAD-ME2* and *GgNAD-ME3* at all measured time points. *GgNAD-ME1* expression peaks midday and is 2.8-fold higher than *GgNAD-ME2* and 7.9-fold higher than *GgNAD-ME3* (8h; $p<0.0001$). This peak *GgNAD-ME1* expression is a significant 5.3-fold increase from early morning (1h) to midday (8h) ($p<0.001$; Fig. 2B). Afterwards, *Gg*NAD-ME1 expression drops to approx. half of the expression strength by the end of the light period (15h, $p<0.001$; Fig. 2B). By the end of the night, expression of *GgNAD-ME1* dropped to a fifth of the peak expression at 8 h, reaching its lowest expression level (Fig. 2B).

Despite the limited comparability of expression strength of genes from different species, it must be noted that all *GgNAD-ME* genes are expressed at higher levels than their paralogous *ThNAD-ME*. *ThNAD-ME1* expression level are similar to that of *ThACT2*, and *ThNAD-ME2* and *ThNAD-ME3* show approx. 20-fold lower expression than this control gene (Fig. 2A). *GgNAD-ME1* is expressed 2- to 10-fold higher than *GgACT2*, while *GgNAD-ME2* range up to 3.7-fold and *GgNAD-ME3* up to 1.3-fold higher expression than *GgACT2* at midday (Fig. 2B).

Overall, expression patterns of all three NAD-ME genes in *T. hassleriana* do not change drastically during a diurnal cycle, which is an expected behavior for a housekeeping function. Our findings did not allow unambiguously identify a concrete $C_4$ *NAD-ME* candidate gene in *G. gynandra*, as all three *NAD-ME* genes are much higher expressed than their homologs in *T. hassleriana*. However, the more pronounced light-dependent expression of *GgNAD-ME1* compared to *ThNAD-ME1* in photosynthetic tissues point out a possible recruitment of *GgNAD-ME1* into the $C_4$ pathway.

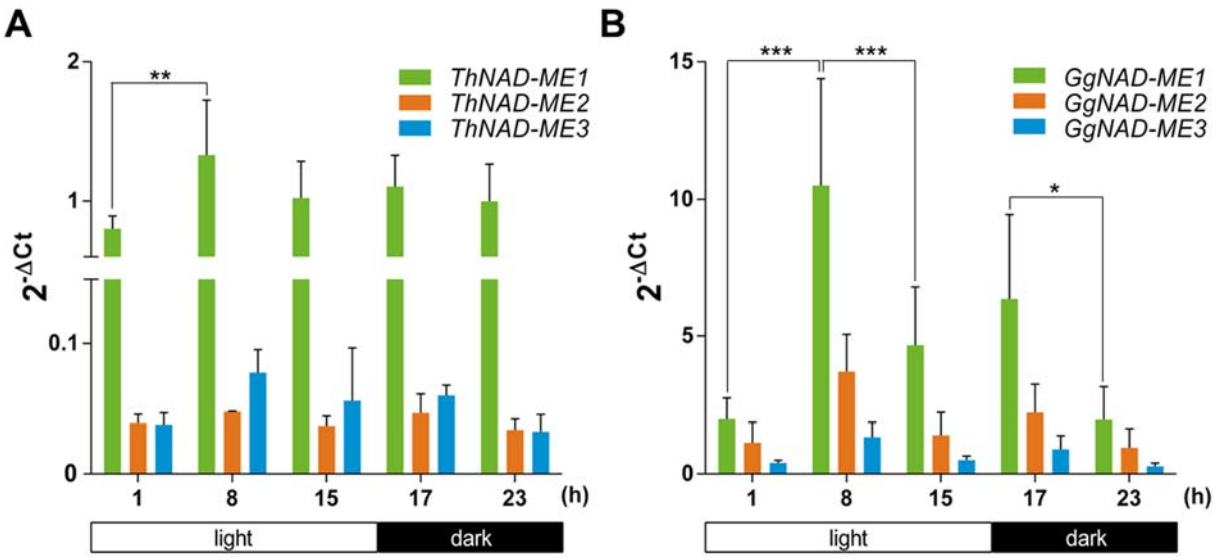



**Figure 2.** Expression levels of *T. hassleriana* (**A**) and *G. gynandra* (**B**) *NAD-ME* genes relative to *ACT2* ($2^{-\Delta Ct}$) in leaves of 5-week-old plants over a full day. Values are mean ± SE (two-way ANOVA, Tukey's post-hoc test $p<0.05$). Asterisks indicate statistical differences: $p<0.05$ (*), $p<0.01$ (**), and $p<0.001$ (***).

### Kinetic properties of recombinant NAD-ME

All three identified NAD-ME transcripts of *T. hassleriana* and their orthologues in *G. gynandra* were cloned into expression vectors for heterologous expression of the predicted mature protein. The recombinant proteins, isolated to homogeneity, presented the expected molecular weights after SDS-PAGE (Suppl. Fig. 2). We determined the kinetic parameters of the single isolated proteins and of all possible combinations of subunits for both Cleome species (14 possible NAD-ME variants). The results of this extensive kinetic analysis are shown in Table 1.

The analysis of the three single recombinant NAD-ME proteins of *T. hassleriana* indicated that all proteins have similar affinities for the substrate malate. *Th*NAD-ME1 (α NAD-ME) has the highest catalytic efficiency due to a high turnover rate. Single recombinant *Th*NAD-ME2 and *Th*NAD-ME3 (β NAD-ME) show an overall similar kinetic behavior with an almost identical turnover rate ($p>0.05$). The analysis of all subunit combinations shows that those including *Th*NAD-ME1 have similar kinetic behavior as *Th*NAD-ME1 ($p>0.05$). Subunit combinations including *Th*NAD-ME3 exhibit a slightly increased affinity towards malate. The kinetic properties of the combination of both β-subunits, *Th*NAD-ME2/*Th*NAD-ME3, are similar to those of the single β-subunits ($p>0.05$), and are 3- to 5-fold lower than those observed with all other subunit combinations ($p<0.0001$). In summary, the kinetic behavior of *Th*NAD-ME

combinations including *Th*NAD-ME1 in their composition is similar to that of the single *Th*NAD-ME1. When assayed alone, both β-subunits and their combination show clearly distinct kinetics, characterized by an up to 10-fold reduction in catalytic efficiency due to low turnover rates.

The analysis of recombinant NAD-ME of $C_4$ *G. gynandra* indicated distinct kinetic behaviors of the single proteins: while *Gg*NAD-ME1 and *Gg*NAD-ME3 are active enzymes with comparable kinetic properties, *Gg*NAD-ME2 is inactive (Table 1). All possible NAD-ME subunit combinations of *Gg*NAD-ME are active and have similar turnover rates, comparable to those of the active homomers ($p>0.05$ for $k_{cat}$). The heteromeric combinations *Gg*NAD-ME1/*Gg*NAD-ME3, *Gg*NAD-ME2/*Gg*NAD-ME3, and *Gg*NAD-ME1/*Gg*NAD-ME2/*Gg*NAD-ME3 show affinities towards malate similar to the homomers *Gg*NAD-ME1 and *Gg*NAD-ME3 ($p>0.05$). Strikingly, *Gg*NAD-ME1/*Gg*NAD-ME2 shows an approx. 10-fold higher affinity towards malate than all other *Gg*NAD-ME possible proteins, rendering an enzyme with high catalytic efficiency (Table 1). Taken all together, the analysis of recombinant NAD-ME proteins indicates that in *G. gynandra* the association of the active subunit *Gg*NAD-ME1 with the inactive subunit *Gg*NAD-ME2 renders a NAD-ME entity with high affinity for the substrate and high catalytic efficiency, expected properties for an enzyme to fulfill a role in a high flux metabolic pathway.

**Table 1.** Kinetic parameters of recombinant NAD-ME from *T. hassleriana* and *G. gynandra*. Kinetic data were fitted by nonlinear regression. Values represent mean ± SE of at least three independent enzyme preparations.

| | $K_m$ malate [mM] | $k_{cat}$ [$s^{-1}$] | $k_{cat}/K_m$ |
|---|---|---|---|
| ***T. hassleriana*** | | | |
| NAD-ME1 | 5.4 ± 1.3 | 29.0 ± 2.1 | 5.4 |
| NAD-ME2 | 9.2 ± 2.3 | 3.1 ± 0.3 | 0.3 |
| NAD-ME3 | 6.8 ± 1.7 | 3.4 ± 0.3 | 0.5 |
| NAD-ME1/NAD-ME2 | 6.6 ± 1.6 | 20.0 ± 1.5 | 3.0 |
| NAD-ME1/NAD-ME3 | 3.4 ± 1.0 | 17.4 ± 1.2 | 5.2 |
| NAD-ME2/NAD-ME3 | 4.1 ± 2.0 | 4.5 ± 0.6 | 1.1 |
| NAD-ME1/NAD-ME2/NAD-ME3 | 3.6 ± 0.8 | 12.9 ± 0.8 | 3.6 |
| ***G. gynandra*** | | | |
| NAD-ME1 | 2.5 ± 2.0 | 2.2 ± 0.9 | 0.9 |
| NAD-ME2 | inactive | inactive | inactive |
| NAD-ME3 | 5.5 ± 0.8 | 6.6 ± 0.3 | 1.2 |
| NAD-ME1/NAD-ME2 | 0.3 ± 0.03 | 2.2 ± 0.5 | 7.3 |
| NAD-ME1/NAD-ME3 | 5.8 ± 0.9 | 4.4 ± 0.2 | 0.8 |
| NAD-ME2/NAD-ME3 | 3.0 ± 0.8 | 4.5 ± 0.3 | 1.5 |
| NAD-ME1/NAD-ME2/NAD-ME3 | 4.0 ± 0.6 | 3.0 ± 0.1 | 1.1 |

**NAD-ME isoforms in photosynthetic and heterotrophic organs of $C_3$ and $C_4$ Cleome**

To determine the composition of NAD-ME present in $C_3$ and $C_4$ Cleome species, we coupled the determination of NAD-ME activity after native PAGE of extracts of leaves and roots of *T. hassleriana* and *G. gynandra* with the identification of NAD-ME proteins in the gel slices through mass spectroscopy (MS). Overall, *in gel* activity assays indicated four protein bands showing NAD-ME activity distributed over species and different tissues (migration levels 1 to 4 in Fig. 3).

Leaves of *G. gynandra* show two main protein bands with NAD-ME activity of very different mobilities (migration level 1 and 4); these bands are not present in the *T. hassleriana* (Fig. 3). The first protein band has a very low mobility (migration level 1, Fig. 3), is visible already after 30 min incubation in the reaction media and is also observed in roots of *G. gynandra,* where it represents the predominant NAD-ME entity (Fig. 3). MS analysis indicated the presence of *Gg*NAD-ME1, *Gg*NAD-ME2, and

*Gg*NAD-ME3 in this protein band. The very low mobility of the protein band suggests that the three *Gg*NAD-ME subunits assemble in a high molecular weight complex. The second protein band found in leaves of $C_4$ *G. gynandra* has very high mobility (migration level 4, Fig. 3) and is not found in roots of *G. gynandra* even after overnight incubation. MS analysis indicated the presence of *Gg*NAD-ME1 and *Gg*NAD-ME2 in this protein band. As the mobility of this protein band is similar to that of the heterodimer NAD-ME1/NAD-ME2 found in *A. thaliana* leaves (Tronconi *et al.*, 2010*a*), we conclude that *Gg*NAD-ME1 and *Gg*NAD-ME2 assemble as heterodimer.

Leaves and roots of *T. hassleriana* possess an protein band with NAD-ME activity that is not found in organs of the $C_4$ species. This NAD-ME entity has a lower mobility than the *A. thaliana* NAD-ME heterodimer (migration level 3, Fig. 3) and the activity appears after long incubation times. MS analysis indicated the presence of all three *Th*NAD-ME subunits (*Th*NAD-ME1, *Th*NAD-ME2 and *Th*NAD-ME3) in this protein band, suggesting the formation of a multimeric complex.

Leaves and roots of G. *gynandra* show an additional band of very low NAD-ME activity and intermediate mobility (migration level 2, Fig. 3). This band is also observed in leaves and roots of $C_3$ *T. hassleriana* (Fig. 3). MS analysis indicated the presence of a NADP-ME (Gg_protein17666, Th_protein09755) in all these protein bands. This suggest that the band at migration level 2 found in all samples most probably represents a promiscuous NADP-ME that can also use NAD as cofactor.

Taking altogether, leaves and roots of the Cleome $C_3$ species *T. hassleriana* possess a unique NAD-ME entity composed of all three subunits, *Th*NAD-ME1, *Th*NAD-ME2 and *Th*NAD-ME3. This entity most probably represents a tetramer, but we cannot exclude that different heteromers could coexist. The Cleome $C_4$ species *G. gynandra* possesses two NAD-ME entities. One, found in both heterotrophic and photosynthetic organs, and thus representing the housekeeping NAD-ME entitiy, is assembled in a high molecular weight heteromer composed of all three subunits, *Gg*NAD-ME1, *Gg*NAD-ME2 and *Gg*NAD-ME3. The second entity, found only in photosynthetic tissue, and thus most probably representing the $C_4$ decarboxylase activity, is assembled as a heterodimer of *Gg*NAD-ME1 and *Gg*NAD-ME2.

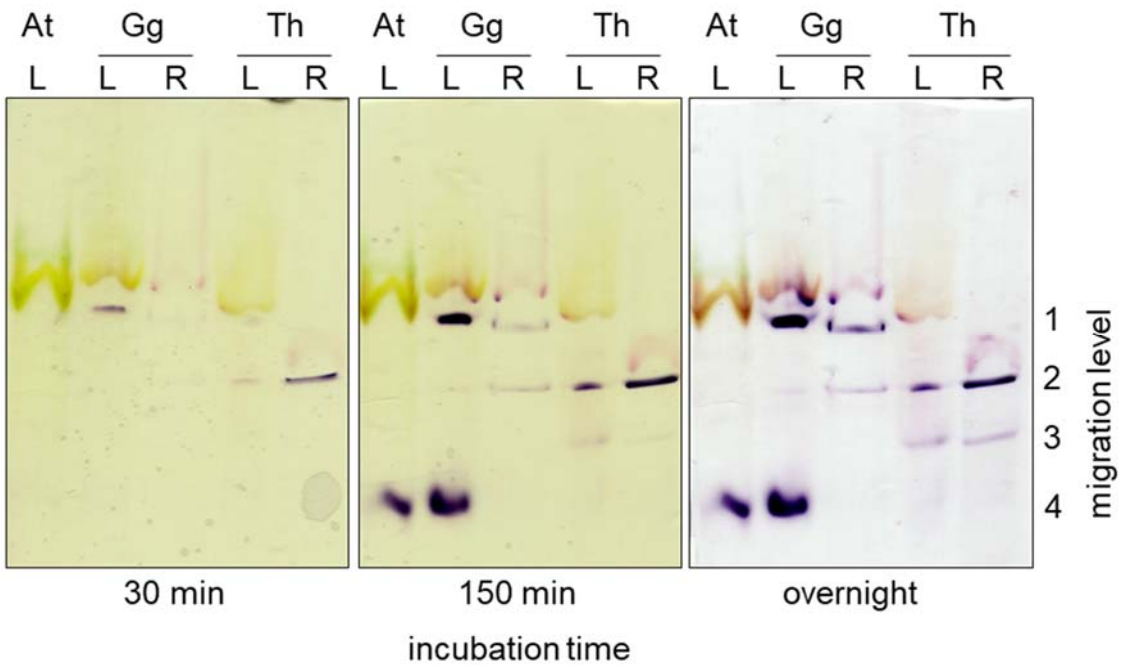



**Figure 3.** NAD-ME activity assays after native PAGE (time series). Non-denaturing PAGE of soluble protein extracts of leaves (L) and roots (R) of *A. thaliana* (At), *G. gynandra* (Gg), and *T. hassleriana* (Th) were incubated for the indicated time periods in the assay medium at pH 6.8. A violet precipitate indicates NAD-ME activity. Migration levels 1 to 4 are indicated on the right.

### *In silico* analysis of β-NAD-ME orthologs

Our analysis of kinetic properties of all possible NAD-ME indicated that *Gg*NAD-ME2 is an inactive protein (Table 1). Moreover, the positioning of *Gg*NAD-ME2 as a sister group to all other β-NAD-ME sequences within the Brassicales (Fig. 1B, Suppl. Fig. 1) is unlikely to reflect the true evolutionary history, but is far more likely a long-branch-attraction artifact caused by an unusual pattern of amino acid substitutions. In order to discover a potential link between those amino acid changes and the lack of activity of *Gg*NAD-ME2, we analyzed the amino acid sequences of closely related β NAD-ME homologs searching for differentially conserved amino acids.

In total, we compared 16 β NAD-ME homologs from species of the same evolutionary lineage (Brassicales). A representative fraction of these 16 sequences is displayed as a multiple sequence alignment in Figure 4, where sequences with similar substitution patterns and thus information redundancy were eliminated. We found 47 amino acid residues within the β NAD-ME mature protein sequence that are changed in *Gg*NAD-ME2 and conserved in all other β NAD-ME sequences, including the paralog *Gg*NAD-ME3. We found 44 amino acid residues 100% conserved in the compared

β NAD-ME sequences (highlighted in blue, Fig. 4), and 3 residues with conserved side chain characteristics (highlighted in orange, Fig. 4).

The 47 differentially conserved amino acids identified are distributed throughout the mature protein sequence, including changes in the malate- (red box, Fig. 4) and NAD-binding (violet box, Fig. 4) domains. The analysis of the amino acid residues of the NAD-ME active site (Chang & Tong, 2003) indicates that two amino acids that participate in the catalytic mechanism, Y132 and D298, are changed in *Gg*NAD-ME2 (Fig. 4, Table 2). These two amino acid substitutions are most probably involved in the observed loss of activity of *Gg*NAD-ME2 as they are strikingly conserved in all other NAD-ME sequences. In addition, the amino acids known to be involved in the metal cofactor coordination (Chang & Tong, 2003) are in all cases conserved (black asterisks, Fig. 4).

**Table 2.** Overview of differentially conserved amino acids of *Gg*NAD-ME2 that are probably involved in substrate coordination and dimer-dimer interaction based on 3D structure modelling and sequence homology to human NAD(P)-ME.

| Location | Brassicales β NAD-ME amino acid | *Gg*NAD-ME2 amino acid | Alignment position |
|---|---|---|---|
| active site | Y | S | 139 |
| active site | D | V | 304 |
| active site | M/L | S | 471 |
| interface | E | K | 261 |
| interface | D | N | 585 |
| interface | I | V | 586 |
| interface | K | T | 596 |
| interface | H | N | 600 |

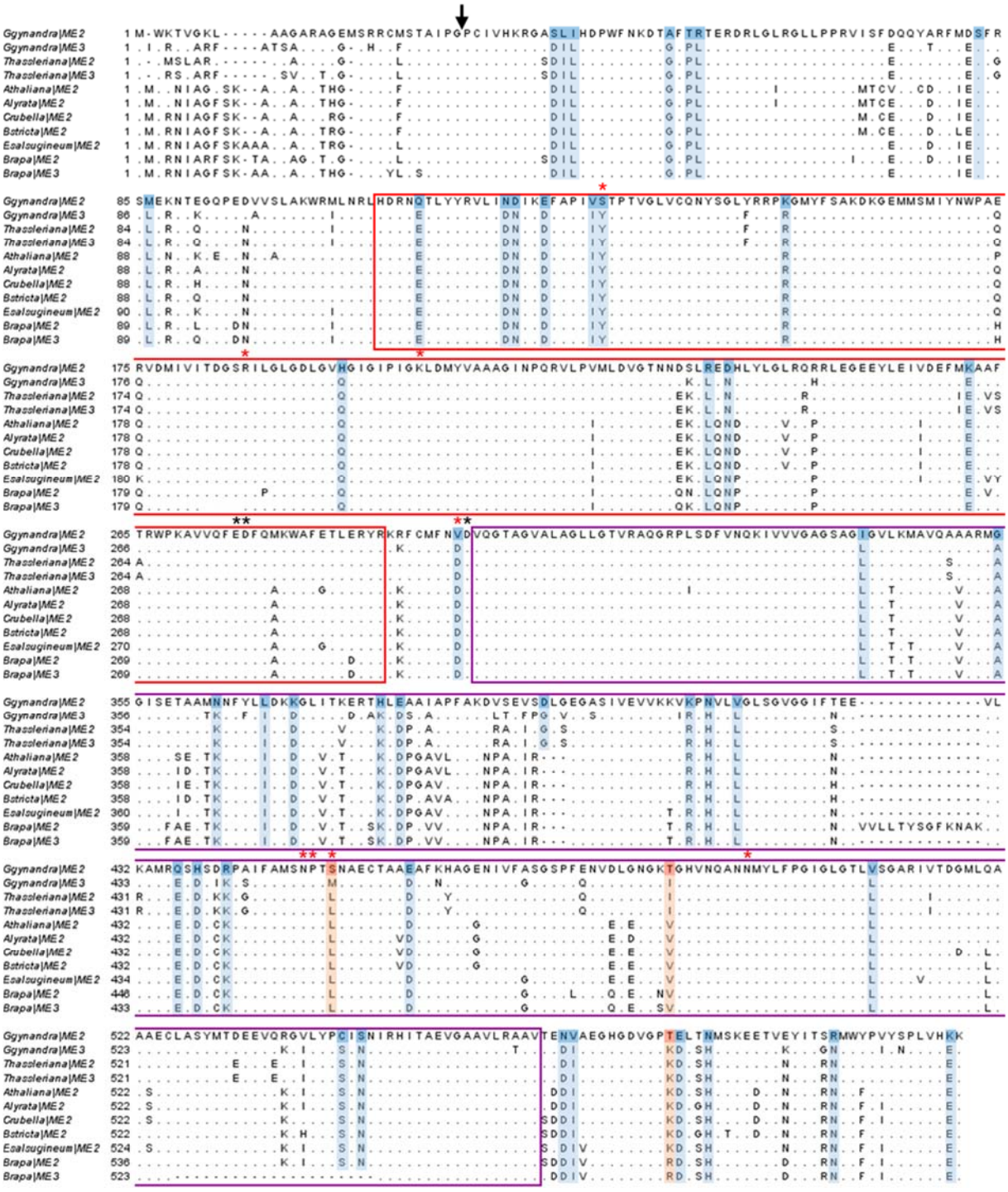



**Figure 4.** Alignment of selected β NAD-ME protein sequences from the Brassicales. *Gg*NAD-ME2 is set as reference; conserved amino acids (AA) are represented by dots, gaps by dashes, and changed AA are shown. Malate and NAD-binding domain are marked in red and violet boxes. Arrow indicates transit peptide cleavage site. Asterisks indicate metal ion (black), malate and NAD coordinating sites (red). Highlighted in are AA differing in *Gg*NAD-ME2 from all other Brassicales (blue: 44 AA 100% conserved in Brassicales β NAD-ME; orange, 3 AA with conserved site chain characteristics).

Human mitochondrial NAD-ME is a tetramer formed by the associations of two dimers that can exists as a mixture of tetramer and dimer at neutral pH (Xu *et al.*, 1999; Chang & Tong, 2003). Our work indicates that *Gg*NAD-ME1/*Gg*NAD-ME2 from $C_4$ Cleome leaves extracts has the same mobility in native PAGE as the *A. thaliana* NAD-ME heterodimer (Fig. 3). Hence, to further predict possible impacts of different amino acid changes found in the sequence of the inactive *Gg*NAD-ME2 we built and analyzed a 3D protein model of heterodimeric *Gg*NAD-ME1/*Gg*NAD-ME2 (Fig. 5A). We based our model on the human mitochondrial NAD(P)-ME crystal structure that is available including different substrates and co-factors (SMTL ID: 1qr6.1 used in Swiss-model, (Biasini *et al.*, 2014). The structure of the active pockets (indicated by asterisks, black for *Gg*NAD-ME1, red for *Gg*NAD-ME2, Fig. 5A) differs due to the amino acid changes in the sequence. One particular changed amino acid at the entrance of the active pocket (S471, Table 2) is normally involved in the coordination of the co-factor binding (Chang & Tong, 2003). A change from hydrophobic side chain characteristic to a polar one may influence the NAD-binding capability of *Gg*NAD-ME2.

Additionally, we found *Gg*NAD-ME in photosynthetic and heterotrophic tissue to be present in a high oligomeric state, probably tetra- or octameric (Fig. 3). To investigate if the changes found in *Gg*NAD-ME2 could influence the oligomerization of this hetero-oligomeric *Gg*NAD-ME, we modeled a tetrameric *Gg*NAD-ME (Fig. 5B, 5C). Several of the amino acid changes to be located at the modeled dimer-dimer interface (Fig. 5C), making them candidates for influencing the transition between dimer and tetrameric state of the enzyme. These changed amino acids (K261, N585, V586, T596, and N600) confer all but one a drastic change of amino acid side chain characteristic (Table 2) and are located outside of the known malate and NAD-binding domains at the C-terminus of the protein.

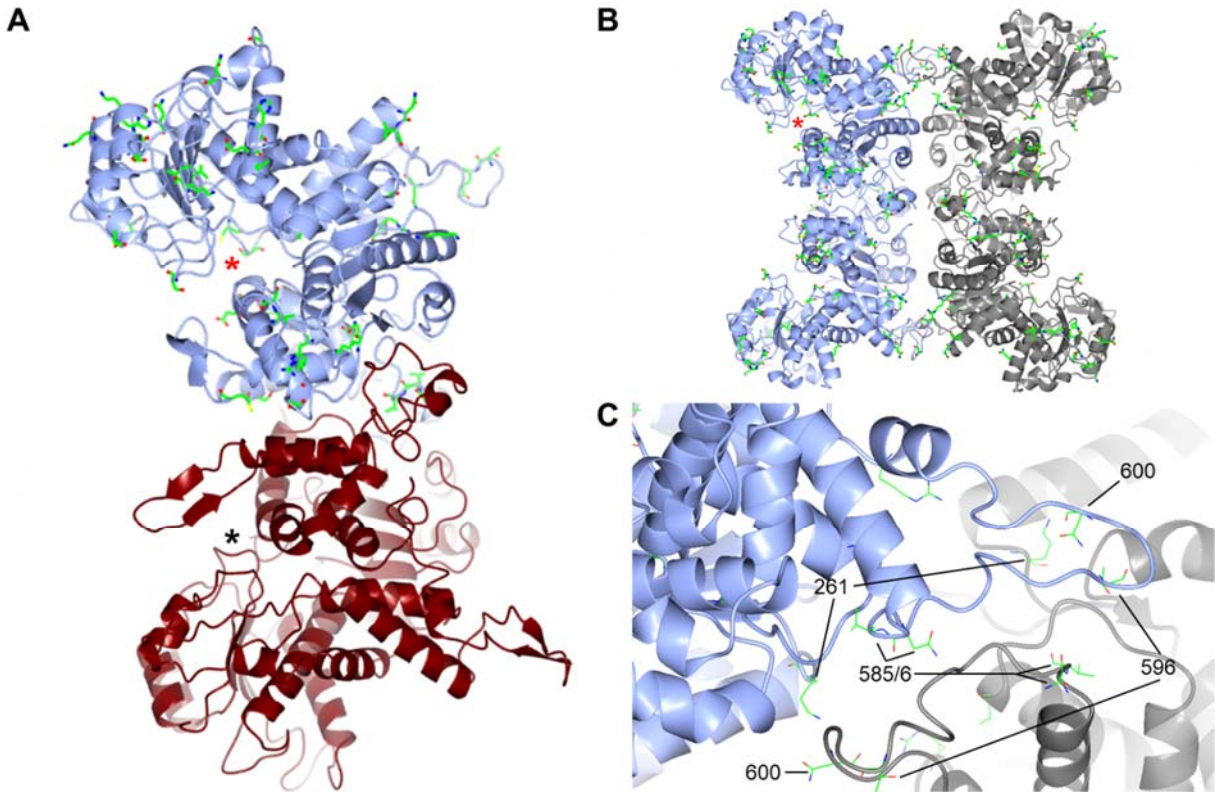



**Figure 5.** Ribbon diagrams of 3D protein structure for two modeled *Gg*NAD-MEs. *Gg*NAD-ME1 monomer is colored in red. Monomers of *Gg*NAD-ME2 are colored in blue and grey. Differentially conserved amino acids are presented in the stick-and-ball configuration. Reactions centers are marked with asterisks. **A.** Heterodimer of *Gg*NAD-ME1/*Gg*NAD-ME2. **B.** Tetramer comprised of two *Gg*NAD-ME2 dimers (blue and grey). **C.** Close-up of the dimer-dimer interface with highlighted and labeled candidate amino acids.

# Discussion

Elucidation of the structural composition and biochemical characteristics of the components of the $C_4$ pathway is not only important to gain knowledge on the evolutionary mechanisms underlying their origin but also because many efforts are currently under way to install characteristics of $C_4$ photosynthesis in leaves of $C_3$ crops. This endeavor can only be fruitful, if it is built on a detailed understanding of the particular characteristics of the components of the $C_4$ pathway. Understanding the recruitment of NAD-ME for the $C_4$ function is intriguing as no $C_4$-specific *NAD-ME* genes were identified so far.

## Evolutionary relationship of NAD-ME proteins

Recently, it was established that the photosynthetic ancestor acquired NAD-ME and NADP-ME in separate evolutionary events (Tronconi *et al.*, 2018). In addition, plant NAD-ME are not evolutionary related with the animal mitochondrial NAD(P)-MEs (Tronconi *et al.*, 2018). Plant mitochondrial NAD-ME is of single gene origin, but we found that at least two genes are present in flowering plants. NAD-ME was likely acquired by horizontal gene transfer from a γ-proteobacteria-like organism to the Archaeplastida ancestor genome at the time point of establishing the evolutionary lineage. The resulting enzyme is speculated to be an enabling event of evolutionary success of the resulting photosynthetic eukaryote (Tronconi *et al.*, 2018).

Our findings on the early evolution of NAD-ME agree with recent findings on the recruitment of NAD-ME as a crucial plant mitochondrial protein. All investigated protein sequences are encoded in the nuclear genome and possess mitochondrial transit peptides. When the *NAD-ME* gene duplication occurred, probably at the basis of the angiosperms, it produced the paralogs coding for α NAD-ME and β NAD-ME. The retained duplication of *NAD-ME* resulting in at least one α NAD-ME and β NAD-ME for all investigated angiosperms is a strong indication for a non-redundant and crucial function in the evolutionary context of NAD-ME in higher plants. The analysis of the evolutionary relationship suggests that the need for a second NAD-ME protein arose latest with the transition to flowering plants (Fig. 1A). In addition, we found that *C. braunii* and *P. patens* retained an evolutionary independent duplication of the *NAD-ME* in their genomes. For *C. braunii*, this might represent a single gene duplication event, as there were no recent whole genome duplication (WGD) event

was detected (Nishiyama *et al.*, 2018). However, for *P. patens* exhibits a retained gene duplication that was caused by the moss specific WGD2 that occurred 38–50 mya. The *P. patens* genes are located on chromosome 7 and 11, which are sister chromosomes produced by the WGD (Lang *et al.*, 2018). Nevertheless, this opens the question what function these new paralogs are fulfilling in the charophytes and mosses.

In *A. thaliana*, as a representative of a higher plant with $C_3$ photosynthesis, the combination of α and β NAD-ME subunits render enzymatic entities that possess specific metabolic functions and regulation through metabolites (Tronconi *et al.*, 2008, 2010*a*, 2010*b*). We found that β NAD-ME duplications have been retained in the genomes of angiosperms more frequently than duplications of α NAD-ME, as this is also the case for the investigated Cleome species (Fig. 1B). Neo-functionalization from gene duplication is a classic driver of protein evolution. The more frequent duplications of the β-subunit a candidate for neo-functionalization. Intriguingly, *Gg*NAD-ME2, one of two β NAD-ME sequences, has accumulated so many amino acid changes that it is placed wrongly at the basis of all Brassicales β NAD-ME (Fig. 1B). *G. gynandra* NAD-ME sequences show generally longer branches in the phylogenetic analysis when compared to its sister species *T. hassleriana*. However, it can be speculated that the coding sequence of *Gg*NAD-ME2 was under positive selection with regards to a possible $C_4$ related function. Gathering more sequences from Cleomaceae should enable a deeper investigation on this hypothesis. This is of special interest, as this might be a shared trait within the $C_4$ Cleomaceae, where *C. angustifolia* and *C. oxalidea* represent two additional $C_4$ origins in this family.

### It is not a single *NAD-ME* gene that was recruited for the $C_4$ function

A straightforward rationale is to assume that if a gene was recruited into the $C_4$ pathway it should be expressed at high levels (Moreno-Villena *et al.*, 2018). A stretch of this hypothesis is the assumption that if one finds all genes for a $C_4$ cycle to be highly expressed in a determined tissue, a functional $C_4$ photosynthesis must be at work (Rangan *et al.*, 2016). The comparative qRT-PCR analysis using $C_3$ and $C_4$ photosynthetic tissue revealed that *NAD-ME* genes are higher expressed in $C_4$ leaves than their homologs in $C_3$ leaves (Fig. 2). Only *GgNAD-ME1* shows significant increase in expression in a light dependent manner, which would be an expected feature of a $C_4$ NAD-ME. Also, the homolog *ThNAD-ME1* showed increased expression during the

first hours in light, albeit in a less strong fashion. Therefore, no clear candidate NAD-ME $C_4$ gene could be identified from the expressional analysis.

Our analysis of heterologously produced NAD-ME proteins indicated that all *T. hassleriana* ($C_3$) NAD-ME proteins are functional enzymes. *Th*NAD-ME proteins are acitiv in vitro as homomeric enzymes as well as in all possible combinations (Table 1). The substrate affinities ($K_m$) of all analyzed NAD-ME entities are in the same range as *A. thaliana* NAD-MEs. This is an expected behavior as both species perform $C_3$ photosynthesis and NAD-ME would be involved in housekeeping functions.

In vitro analysis of *G. gynandra* NAD-ME indicated that only *Gg*NAD-ME1 and *Gg*NAD-ME3 are active enzymes in a homomeric state. *Gg*NAD-ME2 is inactive as a homomeric enzyme, but when combined with the other subunits (*Gg*NAD-ME1 and/or *Gg*NAD-ME3) the properties of the enzymatic entities are changed (Table 1). Interestingly, the *Gg*NAD-ME1/*Gg*NAD-ME2 heteromer shows a much higher affinity towards malate than all other possible *Gg*NAD-ME entities, rendering an enzyme with high catalytic efficiency. When considering the $C_4$ function of NAD-ME in BSC mitochondria, a high catalytic efficiency of the enzyme would be expected. Therefore, our in vitro analysis points out the *Gg*NAD-ME1/*Gg*NAD-ME2 heteromer as a candidate for the $C_4$ function.

**Two NAD-ME entities in $C_4$ photosynthetic tissue**

Albeit some NAD-ME have been described from biochemical data as being homomers from α NAD-ME subunits in some plant species, we found both α and β NAD-ME sequences in all plant genomes and usually heteromeric NAD-ME is found in plants. This also led to the hypothesis, that NAD-ME being an α-β-heteromer is a plant-specific trait and recently it has been postulated that co-evolution of amino acids of these distinct proteins is likely (Tronconi *et al.*, 2018). Our finding that all genomes of flowering plants have retained both NAD-ME genes agrees with this hypothesis.

Our analysis of NAD-ME in $C_3$ and $C_4$ Cleome tissues indicated the presence of NAD-ME entities formed by combinations of subunits of at least one α NAD-ME and one β NAD-ME (Fig. 3). In photosynthetic and heterotrophic tissues of the $C_3$ species *T. hassleriana*, *Th*NAD-ME is present as unique entity composed of *Th*NAD-ME1/*Th*NAD-ME2/*Th*NAD-ME3 (Fig. 3, migration level 3). This is consistent with the idea that NAD-ME fulfills housekeeping functions in every cell in $C_3$ plant

tissues. A similar hetero-oligomeric NAD-ME was identified in photosynthetic and heterotrophic tissues of *G. gynandra,* consisting of *Gg*NAD-ME1/*Gg*NAD-ME2/*Gg*NAD-ME3 (Fig. 3, migration level 1). This NAD-ME entity has a comparable kinetic behavior to that of the NAD-ME found in $C_3$ Cleome. Thus, we propose that this hetero-oligomer (probably a tetra- or octamer) fulfills NAD-ME housekeeping functions in all *G. gynandra* tissues.

Additionally, and uniquely, $C_4$ leaves possess a second NAD-ME entity (Fig. 3, migration level 4). This *Gg*NAD-ME consists of *Gg*NAD-ME1 and *Gg*NAD-ME2 and migrates in a native PAGE together with the major *At*NAD-ME (Fig. 3, level 4), indicating that it is most probable a dimer comprising one *Gg*NAD-ME1 and *Gg*NAD-ME2. In addition, this *Gg*NAD-ME entity possesses kinetic features (Table 1) that allow us to postulate that the heterodimer *Gg*NAD-ME1/*Gg*NAD-ME2 functions as a $C_4$ decarboxylase in *G. gynandra*.

**Duplication of the β-subunit in Cleome aided the evolution of $C_4$ NAD-ME**

We found that in $C_4$ Cleome photosynthetic tissue two distinct NAD-ME entities exist simultaneously. This gives rise to two possible hypotheses regarding the cellular location in the leaf of $C_4$ Cleome (Fig. 6). Either, housekeeping NAD-ME and $C_4$ NAD-ME are separated by cell type as many of the C4 pathway enzymes are (hypothesis A), or the housekeeping NAD-ME is found in every cell type and only $C_4$ NAD-ME is exclusively located in BSCs (hypothesis B).

Expression of *GgNAD-ME1* and *GgNAD-ME2* has been mapped via promoter GUS-staining to *G. gynandra* BSCs in leaf tissue (Brown et al 2011). In that work, *GgNAD-ME3* was not investigated. We argue that the *Gg*NAD-ME1/*Gg*NAD-ME2 dimer is a BSC specific NAD-ME and the expression of the genes is probably very strong in BSC. The presence of the hetero-oligomeric *Gg*NAD-ME1/*Gg*NAD-ME2/*Gg*NAD-ME3 complex and thus the expression of these genes, performing the housekeeping function in the adjacent mesophyll cells in $C_4$ Cleome, could have been overlooked using GUS-staining. If the two NAD-ME entities in $C_4$ Cleome tissues are present within the same cell, in the same mitochondrion at the same time or if they are spatially separated remains to be determined (Fig. 6, A and B).

Nevertheless, both *Gg*NAD-ME entities are formed by the same proteins, in one case assembled into a heterodimer, and in the other as an oligomer of three subunits. One of this subunits, *Gg*NAD-ME2 showed no NAD-ME activity when assayed as a homomeric enzyme (Table1) and accumulated 47 amino acids changes throughout its mature protein sequence that are differentially conserved when compared to other closely related β NAD-ME (Fig. 4). Remarkably, those 47 amino acids are also in the conserved "$C_3$ state" in *Gg*NAD-ME3. We thus propose that these amino acid residue changes in *Gg*NAD-ME2 form the basis for the observed different kinetic properties. The change of two amino acids involved in malate coordination (S139, V304) in the binding site and the change of one co-factor coordinating amino acid (S471) at the entrance of the active site might be responsible for the catalytic inactivity of *Gg*NAD-ME2 (Table 1; Table 2).

We modeled the 3D structure of *Gg*NAD-ME2 (Fig. 5) and identified five of the 47 differentially conserved amino acids as candidates for a regulation of oligomerization (Table 2). These five amino acids are located at the dimer-dimer interface of the modeled tetrameric *Gg*NAD-ME2 (Fig. 5B). While a functional *Gg*NAD-ME would not be comprised solely from *Gg*NAD-ME2, we propose that this β NAD-ME subunit is present as part of a heteromer regulating the transition between a hypothesized $C_4$ NAD-ME dimer and a multimeric housekeeping NAD-ME. Interestingly, the newly emerged amino acids at the freely accessible dimer-dimer interface are ones that are prone to post-translational modification like phosphorylation (threonine) or display change side chain polarity (lysine, asparagine) that can influence interaction strengths. Oligomer transition has been shown to influence kinetic properties in potato NAD-ME (Grover & Wedding, 1982) and has been discussed as a basis for function in an NAD-ME CAM plant, *K. fedtschenkoi* (Cook *et al.*, 1995). In addition to a potential role in the oligomerization status, *Gg*NAD-ME2 seems to increase the affinity of the α-β heteromer towards malate and increases the catalytic efficiency, while in combination with *Gg*NAD-ME3 (β-β heteromer), only marginal effects on $K_m$ and $k_{cat}$ were found. This strengthens the hypothesis of *Gg*NAD-ME2 being a key player in building $C_4$ NAD-ME and potentially regulating the presence of both housekeeping NAD-ME and $C_4$ NAD-ME if these two entities are present in the same cell (Fig. 6, hypothesis B).

Contrasting, *Gg*NAD-ME3 containing homo- and heteromers, including the postulated housekeeping *Gg*NAD-ME1/*Gg*NAD-ME2/*Gg*NAD-ME3 hetero-oligomer, present

kinetic properties close to $C_3$ Cleome hetero-oligomer (Table 1). This underlines the differences between the two β *Gg*NAD-MEs found and points to a specialized role of *Gg*NAD-ME2 in heteromers that are found in $C_4$ photosynthetic tissue. We thus propose that the retained β NAD-ME duplication in the genus Cleome was a potentiating event for the evolution of $C_4$ photosynthesis, enabling the accumulation of multiple amino acid changes in one copy β NAD-ME of the $C_4$ species *G. gynandra*.

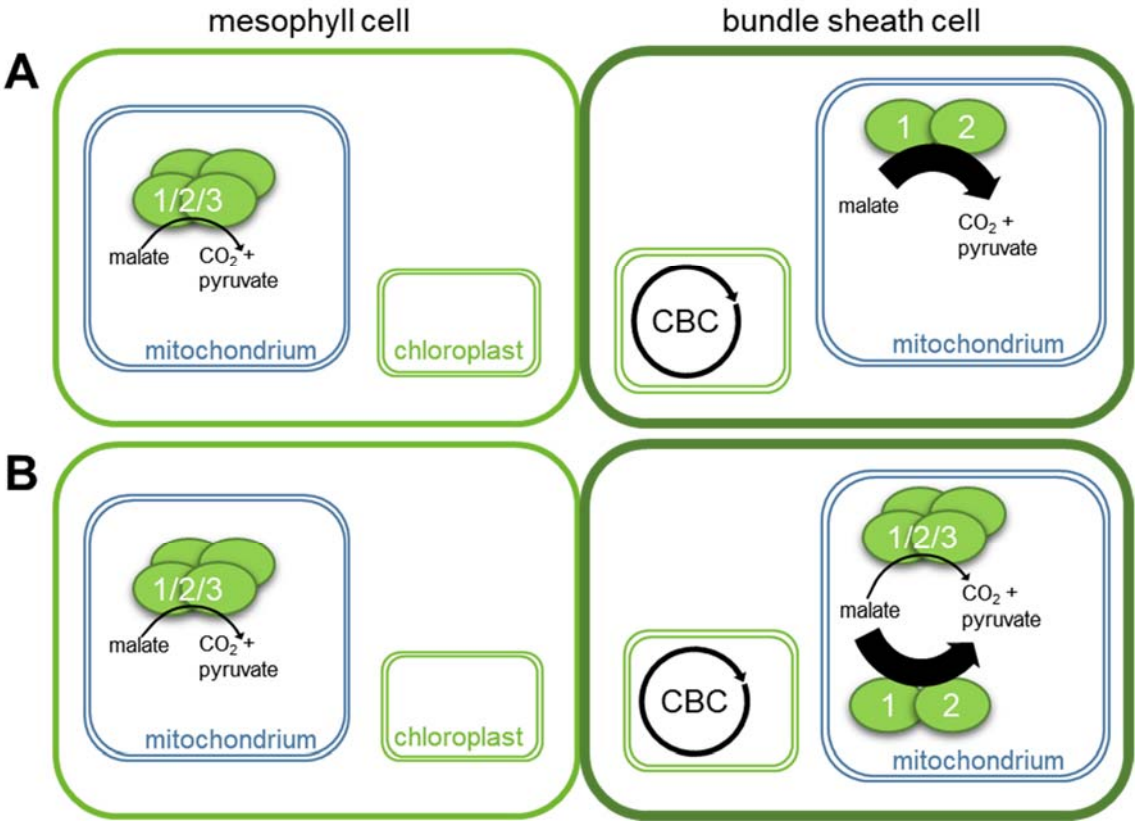



**Figure 6.** Schematic representation of *G. gynandra* NAD-ME complexes and their presence in mesophyll (M) and bundle sheath cells (BSC). Dimeric *Gg*NAD-ME1/*Gg*NAD-ME2 is present in the BSC as part of the $C_4$ pathway. CBC, Calvin Benson cycle. Additionally, multimeric *Gg*NAD-ME comprised of all three subunits (GgNAD-ME1, -ME2, and -ME3) is present in M cells only (**A**) or both in M and BSC (**B**) fulfilling the NAD-ME housekeeping function.

# Materials and Methods

## Phylogenetic analysis and sequence retrieval

NAD-ME sequences from species whose genome assemblies are available on Phytozome (phytozome.jgi.doe.gov) were retrieved using on site blast search service and bulk download. Sequences from *Cleome* species were originally acquired from transcriptome data (Külahoglu *et al.*, 2014), verified and corrected by cDNA based sequencing. Sequences for *C. braunii, A. rosea*, and *A. prostrata* were extracted from data accompanying respective genome publication articles (Nishiyama *et al.*, 2018). All sequences were manually checked for correct translation start and mitochondrial localization was verified using Target P (Emanuelsson *et al.*, 2000).

Full length dataset multiple sequence alignments (MSAs) were computed using the MAFFT algorithm (v7.402) with the iterative refinement method L-INS-i (Katoh *et al.*, 2005; Katoh & Standley, 2013). The best amino acid substitution model based on each MSA was estimated using MEGA X (v.10.0.4) (Kumar *et al.*, 2018). Maximum likelihood (ML) trees were computed using the amino acid substitution model JJT+F, allowing for invariant sites and discrete gamma distributed categories (n = 5) (Jones *et al.*, 1992). Amino acid subsets were defined by 60 % threshold for deletion of alignment columns, meaning only 40% gaps were allowed in a column to be included in the final dataset. This resulted in n = 605 (full set of sequences) and n = 608 (for the Brassicales subset) amino acid positions to be included. Initial trees for the ML were determined with default parameters and an exhaustive ML space was searched (branch swap filter: very strong).

## Plant growth conditions

*Gynandropsis gynandra* and *Tarenaya hassleriana* were grown under greenhouse conditions in commercially available soil mixed with Cocopor® (Stender, Schermbeck, Germany). *G. gynandra* was germinated in the dark for 3-5 days, while *T. hassleriana* needed approx. 14 days for germination under long day (16h/8h) light regime. Natural light was supplemented with regular filament lamps mounted 1.2 m above the surface ensuring a minimum of 80 µmol $m^{-2}$ $s^{-1}$ PAR. Seedlings were transferred from germination pots to single clay pots after full formation of the cotyledons. Plants were watered from above as needed and single pot soil was premixed with 3 g $L^{-1}$ Osmocote

(Scotts Deutschland GmbH, Nordhorn, Germany) as fertilizer. *A. thaliana* (ecotype Columbia-0) was grown in soil as described before (Hüdig *et al.*, 2015).

**Transcriptional analysis of NAD-ME genes**

Total RNA from samples were isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) from pooled (n = 5) plant leaf material of *G. gynandra* and *T. hassleriana*. Genomic DNA was removed by DNAse treatment (Ambion Inc., Austin, USA), according to the manufacturer's instructions. cDNA synthesis was performed using RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific, Darmstadt, Germany) with oligodT primers according to the manufacturer's instructions. Primer for quantitative real-time polymerase chain reaction (qRT-PCR) were designed to amplify a PCR product of 130-210 bp length in the region 450-750 bp from the 3'-end of the coding sequence of the mRNA transcript. The respective Actin2 homolog was chosen from nearest orthologs to *A. thaliana* Actin2. Primer efficiency was determined with regards to primer concentration and annealing temperature, and was used to normalize $C_t$ values. Primer specificity was ensured using cloned coding sequences from plasmids as test templates and agarose gel visualisation of the respective final qRT-PCR products. qRT-PCR was run on StepOne Plus (Thermo Fisher Scientific, Darmstadt, Germany) equipped with StepOne Software v2.2.2 using KAPA SYBR FAST qPCR Kit Master Mix (2X) ABI Prism (Kapabiosystems, Boston, USA) in a 15 µL reaction. qRT-PCR amplification was performed at: 95 °C 3 min, 40 cycles at 95 °C for 3 s and 65 °C for 20 s, followed by a standard melting curve.

**Cloning**

Coding sequences of the NAD-ME subunits of *G. gynandra* and *T. hassleriana* were amplified from the respective leaf tissue cDNA using Phusion Polymerase (Thermo Fisher Scientific, Darmstadt, Germany) and specific primers (Suppl. Table 1). The amplified fragments were sub-cloned into pCR-TOPO-BluntII (Thermo Fisher Scientific, Darmstadt, Germany) and sequenced using EZ-seq sequencing services (Marcorgen Europe, Amsterdam, The Netherlands). The generated TOPO vectors were used as source for generation of cohesive-end insert fragments for restriction site cloning or as templates for further PCR-fragment constructions according to Gibson (2011). The vector pET16b (Merck, Darmstadt, Germany) was used for heterologous expression in *E. coli*. When using classic restriction enzyme cloning, pET16b vector

was linearized with restriction enzymes (*Nde*I/*Xho*I/*BamH*I) as well as the respective PCR products to generate the cohesive ends and circular plasmids were reconstituted with T4 Ligase (Thermo Fisher Scientific, Darmstadt, Germany). If Gibson assembly was used, the primers were designed with a 20 bp overlap with the destination vector. Purified Gibson-assembly PCR-fragments of NAD-ME coding sequences were cloned into pET16b linearized with *BamH*I. Gibson isothermal assembly was performed according to Gibson (2011) with reagents received from Thermo Fisher Scientific (Darmstadt, Germany) and New England Biolabs (Ipswich, USA). Final plasmids were sequenced and amplified using *E. coli* DH5α cells (Thermo Fisher Scientific, Darmstadt, Germany).

**Heterologous expression and purification of recombinant NAD-ME**

Heterologous protein production was performed with the T7-polyperase IPTG-inducible expression system. For production of HIS:GgNAD-ME1 and HIS:ThNAD-ME1, *E. coli* BL21 (DE3) cells (Merck, Darmstadt, Germany) (from now on denoted as protocol 1, P1) and for production of HIS:GgNAD-ME2, HIS:GgNAD-ME3, HIS:ThNAD-ME2, and HIS:ThNAD-ME3, *E. coli* ArcticExpress (DE3) cells (Agilent technologies, Santa Clara, USA) (from now on denoted as protocol 2, P2) were used. Chemically competent *E. coli* cells were transformed with 50 ng of the expression vector and grown on LB agar plates with antibiotics according to the manufacturer's instructions. For protein expression, 400 mL liquid cell cultures containing 100 µg/mL ampicillin (P1) or 100 µg/mL ampicillin and 20 µg/mL gentamicin (P2) were inoculated with freshly grown over-night culture, grown at 37 °C for 2.5 h to an $OD_{600} \approx 0.5$ (P1) or at 37 °C for 3 h to an $OD_{600} \approx 0.8$ (P2). Protein expression was induced with 1 mM IPTG. Cells were harvested after 20 h growth at 37°C (P1) or 3 days grown at 12 °C (P2) by centrifugation at 4 °C/6000 g, and stored at -20 °C until use for a maximum of three month.

Purification of HIS-tagged proteins was performed using the immobilized metal ion affinity chromatography principle. *E. coli* cells were resuspended in freshly prepared ice cold lysation buffer (0.2 M NaCl, 20 mM Tris-HCl pH 8.0, 5 mM imidazole, 2 mM phenylmethylsulfonyl fluoride (PMSF), 1 spatula tip lysozyme), incubated for 10 min on ice and sonified for 4 min in 30 s intervals. After centrifugation (15 min, 15000 g, 4°C), supernatant was loaded on 900 µL Ni-NTA agarose (Qiagen, Hilden, Germany). Three washing steps with incrementing imidazole concentration were performed

(0.2 M NaCl, 20 mM Tris-HCl pH 8.0, and 5, 40, 80 mM (P1) or 5, 40, 60 mM (P2) imidazole) at 4°C with 10 column volumes each until purity of the respective protein was achieved. Proteins were eluted in 2 mL elution buffer (0.2 M NaCl, 20 mM Tris-HCl pH 8.0, 0.5 M imidazole) and buffer exchange to 10 mM MOPS-NaOH (pH 6.8) was performed including concentration and further purification of protein with Amicon Ultra 0.5 mL 50K centrifugal devices (Merck Millipore, Darmstadt, Germany).

**Protein quantification and gel electrophoresis procedures**

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to monitor protein purification according to Laemmli (1970). Native PAGE was used to separate native protein complexes from plant tissue (native PAGE) by omitting β-mercaptoethanol and SDS from any buffer used for sample loading, pouring or running PAGE gels. Proteins in gels were visualized using Coomassie® Brilliant Blue staining procedure. Protein concentration was determined by a simplified amido black 10B precipitation method (Schaffner & Weissmann, 1973): the aqueous protein sample (100 µL) was precipitated directly in a 1.5 mL reaction tube with 400 µL staining solution (90 % (v/v) methanol, 10 % (v/v) acetic acid, 0.01 % (w/v) amido black 10B) by centrifugation (20000 g, 12 min, room temperature (RT)). The precipitate was washed once (1 mL 90 % (v/v) methanol, 10 % (v/v) acetic acid) followed by centrifugation (20000 g, 10 min, RT), air dried and resuspended in 1 mL 0.2 M NaOH. Absorption was measured at 615 nm and known bovine serum albumin concentrations were used as standard.

**Preparation of protein extracts for native PAGE, *in-gel* NAD-ME activity assay and immunological detection**

Plant tissue was freshly harvested from greenhouse grown *G. gynandra*, *T. hassleriana*, and *A. thaliana* and ground to fine powder in liquid nitrogen. Extraction of soluble proteins was performed by adding 3 µL per mg fresh weight ice-cold extraction buffer (0.1 M Tris-HCl pH 8.0, 0.1 M NaCl, 0.5 % (v/v) Triton X-100, 2 mM PMSF, 10 mM dithiothreitol (DTT), 0.1% (w/v) polyvinylpyrrolidone 40), followed by vigorous vortexing and incubation for 20 min on ice. Soluble protein was separated by centrifugation (10000 g, 10 min, 4 °C) and supernatant was transferred to a new tube. Proteins were separated in native PAGE run at 100 V, 4°C until the dye front reached

the bottom of the gel. After electrophoresis, gels were assayed for NAD-ME activity or used for Western blotting.

For NAD-ME activity assay, gels were first incubated in 50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS)-NaOH (pH 6.8) for 30 min. Rebuffered native PAGE gels were incubated in 50 mM MOPS-NaOH (pH 6.8), 0.05 % (w/v) nitro blue tetrazolium (NBT), 150 µM PMS, 10 mM $MnCl_2$, 5 mM $NAD^+$ and either 20 mM, 100 mM or no malate. Gels were documented after initial clearance in $ddH_2O$ and gel parts with activity stain were excised from the gel and processed for protein identification by mass spectrometry.

**Mass spectrometry**

Slices of PAGE samples for mass spectrometry were processed as described in (Schmitz *et al.*, 2017*a*). Acquired spectra were matched against custom-made *G. gynandra* and *T. hassleriana* protein databases.

**Kinetic analysis**

Purified recombinant NAD-ME was assayed for malic enzyme activity using the following reaction mixture: 50 mM MOPS-NaOH (pH 6.8), 10 mM $MnCl_2$, 5 mM NAD, 0.2 µg-2 µg protein and varying concentration of L-malate (0-50 mM) in a total volume of 200 µL in a 96 flat bottom well plate (Greiner Bio-One, Solingen, Germany). Only the combination of *Gg*NAD-ME1/*Gg*NAD-ME2 were measured at pH 6.3. The reaction was followed spectrophotometrically at 340 nm using Synergy HT platereader (BioTek Instruments, Winooski, USA) equipped with Gen5 version 2.07.17 software. All reactions were performed with technical triplicates. Initial velocity was determined by slope calculation of the linear regression of five linear time points using a custom made R script run in R version 3.4.2 (R Core Team, 2017). An extinction coefficient ($\varepsilon$) of 6.22 $mM^{-1}$ $cm^{-1}$ for NADH at 340 nm was used in the calculations. Means of technical triplicates were treated as biological replicates for subsequent calculation of kinetic parameters with GraphPad Prism (v. 6.01) software. All kinetic parameters were calculated at least with four batches of independently purified proteins.

**Statistical analyses**

Statistical analysis for all experiments was performed using GraphPad Prism (v. 6.01) software using a two-way analysis of variance (ANOVA; with a p value of $p<0.05$) and Tukey correction for post-hoc multiple comparisons.

## Acknowledgements

We thank Esther Sundermann (Institut of Computational Cell Biology, Heinrich Heine University Düsseldorf) for providing a custom made R script. This work was supported by grants of the Deutsche Forschungsgemeinschaft EXC1028 and MA2379/18-1 to VGM.

## References

Aubry S, Brown NJ and Hibberd JM. (2011) The role of proteins in $C_3$ plants prior to their recruitment into the $C_4$ pathway. *Journal of Experimental Botany* **62**, 3049–3059.

Barker MS, Vogel H and Schranz ME. (2009) Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* **1**, 391–399.

Bayat S, Schranz ME, Roalson EH and Hall JC. (2018) Lessons from Cleomaceae, the Sister of Crucifers. *Trends in Plant Science* **23**, 808–821.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L and Schwede T. (2014) SWISS-MODEL : modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**, 252–258.

Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Mass J, Lercher MJ, Westhoff P, Hibberd JM and Weber APM. (2011) An mRNA blueprint for $C_4$ photosynthesis derived from comparative transcriptomics of closely related $C_3$ and $C_4$ species. *Plant Physiology* **155**, 142–156.

Chang GG and Tong L. (2003) Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry* **42**, 12721–12733.

Christin P-A, Boxall SF, Gregory R, Edwards EJ, Hartwell J and Osborne CP. (2013) Parallel recruitment of multiple genes into $C_4$ photosynthesis. *Genome Biology and Evolution* **5**, 2174–2187.

Christopher JT and Holtum J. (1996) Patterns of carbon partitioning in leaves of crassulacean acid metabolism species during deacidification. *Plant Physiology* **112**, 393–399.

Cook RM, Lindsay JG, Wilkins MB and Nimmo HG. (1995) Decarboxylation of malate in the crassulacean acid metabolism plant *Bryophyllum* (*Kalanchoë*) *fedtschenkoi*. *Plant Physiology* **109**, 1301–1307.

Dever L V, Pearson M, Ireland RJ, Leegood RC and Lea PJ. (1998) The isolation and characterisation of a mutant of the $C_4$ plant *Amaranthus edulis* deficient in NAD-malic enzyme activity. *Planta* **206**, 649–656.

Drincovich MF, Lara M V, Andreo CS and Maurino VG. (2011) $C_4$ decarboxylases: different solutions for the same biochemical problem, the provision of $CO_2$ to Rubisco in the bundle sheath cells. *In $C_4$ Photosynthesis and Related $CO_2$ Concentrating Mechanisms, Vol. 32,* 277–300.

Edwards EJ, Osborne CP, Strömberg CAE, Smith SA and $C_4$ Grasses Consortium. (2010) The origins of $C_4$ Grasslands: integrating evolutionary and ecosystem science. *Science* **328**, 587–591.

Emanuelsson O, Nielsen H, Brunak S and Von Heijne G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005–1016.

Feodorova TA, Voznesenskaya EV, Edwards GE and Roalson EH. (2010) Biogeographic Patterns of Diversification and the Origins of $C_4$ in *Cleome* (Cleomaceae). *Systematic Botany* **35**, 811–826.

Furbank RT. (2011) Evolution of the C4 photosynthetic mechanism: are there really three $C_4$ acid decarboxylation types? *Journal of Experimental Botany* **62**, 3103–3108.

Furbank RT and Hatch MD. (1987) Mechanism of $C_4$ photosynthesis. *Plant Physiology* **85**, 958–964.

Gibson DG. (2011) Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology* **498**, 349–361.

Grover SD and Wedding RT. (1982) Kinetic ramifications of the association-dissociation behavior of NAD malic enzyme. *Plant Physiology* **70**, 1169–1172.

Grover SD, Canellas PF and Wedding RT. (1981) Purification of NAD malic enzyme from potato and investigation of some physical and kinetic properties. *Archives of Biochemistry and Biophysics* **209**, 396–407.

Hall JC, Sytsma KJ and Iltis HH. (2002) Phylogeny of Capparaceae and Brassicacecae based on chloroplast sequence data. *American Journal of Botany* **89**, 1826–1842.

Hüdig M, Maier A, Scherrers I, Seidel L, Jansen EEW, Mettler-Altmann T, Engqvist MKM and Maurino VG. (2015) Plants possess a cyclic mitochondrial metabolic pathway similar to the mammalian metabolic repair mechanism involving malate dehydrogenase and L-2-hydroxyglutarate dehydrogenase. *Plant and Cell Physiology* **56**, 1820–1830.

Jones DT, Taylor WR and Thornton JM. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**, 275–282.

Katoh K, Kuma KI, Toh H and Miyata T. (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**, 511–518.

Katoh K and Standley DM. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.

Koteyeva NK, Voznesenskaya EV, Roalson EH and Edwards GE. (2011) Diversity in forms of $C_4$ in the genus *Cleome* (Cleomaceae). *Annals of Botany* **107**, 269–283.

Külahoglu C, Denton AK, Sommer M, Maß J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo E, Buell CR, Simon R, De Veylder L, Bräutigam A and Weber APM. (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae $C_3$ and $C_4$ plant species. *Plant Cell* **26**, 3243–3260.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549.

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

Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB, Piednoel M, Gundlach H, Van Bel M, Meyberg R, Vives C, Morata J, Symeonidi A, Hiss M, Muchero W, Kamisugi Y, Saleh O, Blanc G, Decker EL, van Gessel N, Grimwood J, Hayes RD, Graham SW, Gunter LE, McDaniel SF, Hoernstein SNW, Larsson A, Li FW, Perroud PF, Phillips J, Ranjan P, Rokhsar DS, Rothfels CJ, Schneider L, Shu S, Stevenson DW, Thümmler F, Tillich M, Villarreal Aguilar JC, Widiez T, Wong GK-S, Wymore A, Zhang Y, Zimmer AD, Quatrano RS, Mayer KFX, Goodstein D, Casacuberta JM, Vandepoele K, Reski R, Cuming AC, Tuskan GA, Maumus F, Salse J, Schmutz J and Rensing SA. (2018) The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant Journal* **93**, 515–533.

Lauterbach M, Schmidt H, Billakurthi K, Hankeln T, Westhoff P, Gowik U and Kadereit G. (2017) *De novo* transcriptome assembly and comparison of $C_3$, $C_3$-$C_4$, and $C_4$ species of tribe Salsoleae (Chenopodiaceae). *Frontiers in Plant Science* **8**.

Long JJ, Wang J-L and Berry JO. (1994) Cloning and analysis of the $C_4$ photosynthetic NAD-dependent malic enzyme of amaranth mitochondria. *The Journal of Biological Chemistry* **269**, 2827–2833.

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

Matsuoka M, Furbank RT, Fukayama H and Miyao M. (2001) Molecular engineering of $C_4$ photosynthesis. *Annual Review Of Plant Physiology and Plant Molecular Biology* **52**, 297–314.

Monson RK. (2003) Gene duplication, neofunctionalization, and the evolution of $C_4$ photosynthesis. *International Journal of Plant Science* **164**, 43–54.

Moreno-Villena JJ, Dunning LT, Osborne CP and Christin PA. (2018) Highly expressed genes are preferentially co-opted for $C_4$ photosynthesis. *Molecular Biology and Evolution* **35**, 94–106.

Murata T, Ohsugi R, Matsuoka M and Nakamoto H. (1989) Purification and characterization of NAD malic enzyme from leaves of *Eleusine coracana* and *Panicum dichotomiflorum*. *Plant Physiology* **89**, 316–324.

Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, Vosolsobě S, Rombauts S, Wilhelmsson PKI, Janitza P, Kern R, Heyl A, Rümpler F, Villalobos LIAC, Clay JM, Skokan R, Toyoda A, Suzuki Y, Kagoshima H, Schijlen E, Tajeshwar N, Catarino B, Hetherington AJ, Saltykova A, Bonnot C, Breuninger H, Symeonidi A, Radhakrishnan G V., Van Nieuwerburgh F, Deforce D, Chang C, Karol KG, Hedrich R, Ulvskov P, Glöckner G, Delwiche CF, Petrášek J, Van de Peer Y, Friml J, Beilby M, Dolan L, Kohara Y, Sugano S, Fujiyama A, Delaux PM, Quint M, Theißen G, Hagemann M, Harholt J, Dunand C, Zachgo S, Langdale J, Maumus F, Van Der Straeten D, Gould SB and Rensing SA. (2018) The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448–464.

Ohsugi R and Murata T. (1980) Leaf anatomy, post-illumination $CO_2$ burst and NAD-malic enzyme activity of *Panicum dichotomiflorum*. *Plant and Cell Physiology* **21**, 1329–1333.

Rangan P, Furtado A and Henry RJ. (2016) New evidence for grain specific $C_4$ photosynthesis in wheat. *Scientific Reports* **6**, 31721.

Sage RF, Christin P-A and Edwards EJ. (2011) The $C_4$ plant lineages of planet Earth. *Journal of Experimental Botany* **62**, 3155–3169.

Saigo M, Bologna FP, Maurino VG, Detarsio E, Andreo CS and Drincovich MF. (2004) Maize recombinant non-$C_4$ NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Molecular Biology* **55**, 97–107.

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

Schmitz J, Srikanth NV, Hüdig M, Poschmann G, Lercher MJ and Maurino VG. (2017) The ancestors of diatoms evolved a unique mitochondrial dehydrogenase to oxidize photorespiratory glycolate. *Photosynthesis Research* **132**, 183–196.

Schranz ME and Mitchell-Olds T. (2006) Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* **18**, 1152–1165.

Tausta SL, Miller Coyle H, Rothermel B, Stiefel V and Nelson T. (2002) Maize $C_4$ and non-$C_4$ NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Molecular Biology* **50**, 635–652.

Team R Core. (2017) R: A language and environment for statistical computing.

Tronconi MA, Andreo CS and Drincovich MF. (2018) Chimeric structure of plant malic enzyme family: different evolutionary scenarios for NAD- and NADP-dependent isoforms. *Frontiers in Plant Science* **9**, 1–15.

Tronconi MA, Maurino VG, Andreo CS and Drincovich MF. (2010a) Three different and tissue-specific NAD-malic enzymes generated by alternative subunit association in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **285**, 11870–11879.

Tronconi MA, Gerrard Wheeler MC, Maurino VG, Drincovich MF and Andreo CS. (2010b) NAD-malic enzymes of *Arabidopsis thaliana* display distinct kinetic mechanisms that support differences in physiological control. *The Biochemical Journal* **430**, 295–303.

Tronconi MA, Wheeler MCG, Martinatto A, Zubimendi JP, Andreo CS and Drincovich MF. (2015) Allosteric substrate inhibition of Arabidopsis NAD-dependent malic enzyme 1 is released by fumarate. *Phytochemistry* **111**, 37–47.

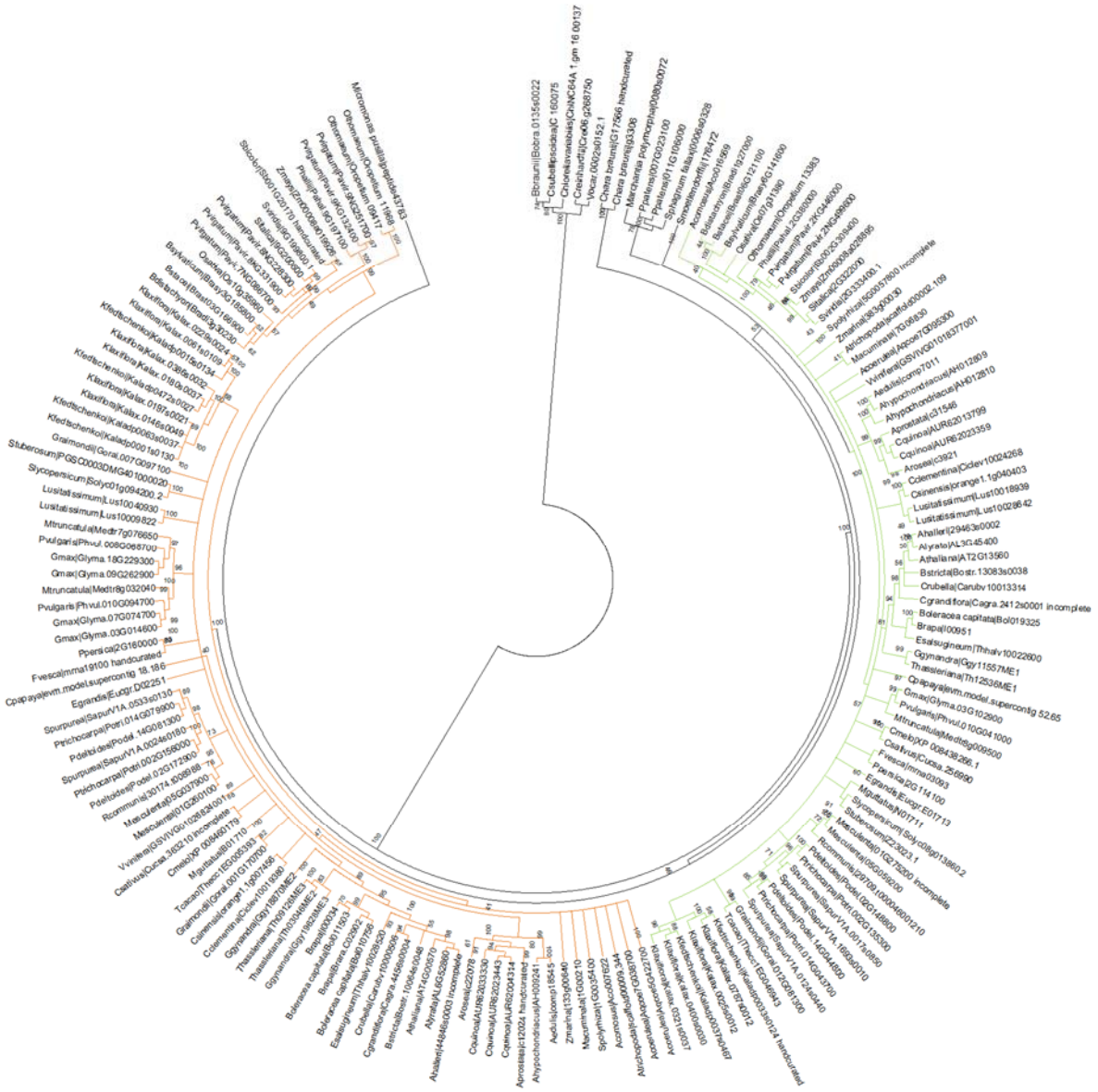
Tronconi MA, Fahnenstich H, Gerrard Wheeler MC, Andreo CS, Flügge U-I, Drincovich MF and Maurino VG. (2008) *Arabidopsis* NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact on nocturnal metabolism. *Plant Physiology* **146**, 1540–1552.

Wang Y, Long SP and Zhu X-G. (2014) Elements required for an efficient NADP-malic enzyme type $C_4$ photosynthesis. *Plant Physiology* **164**, 2231–2246.

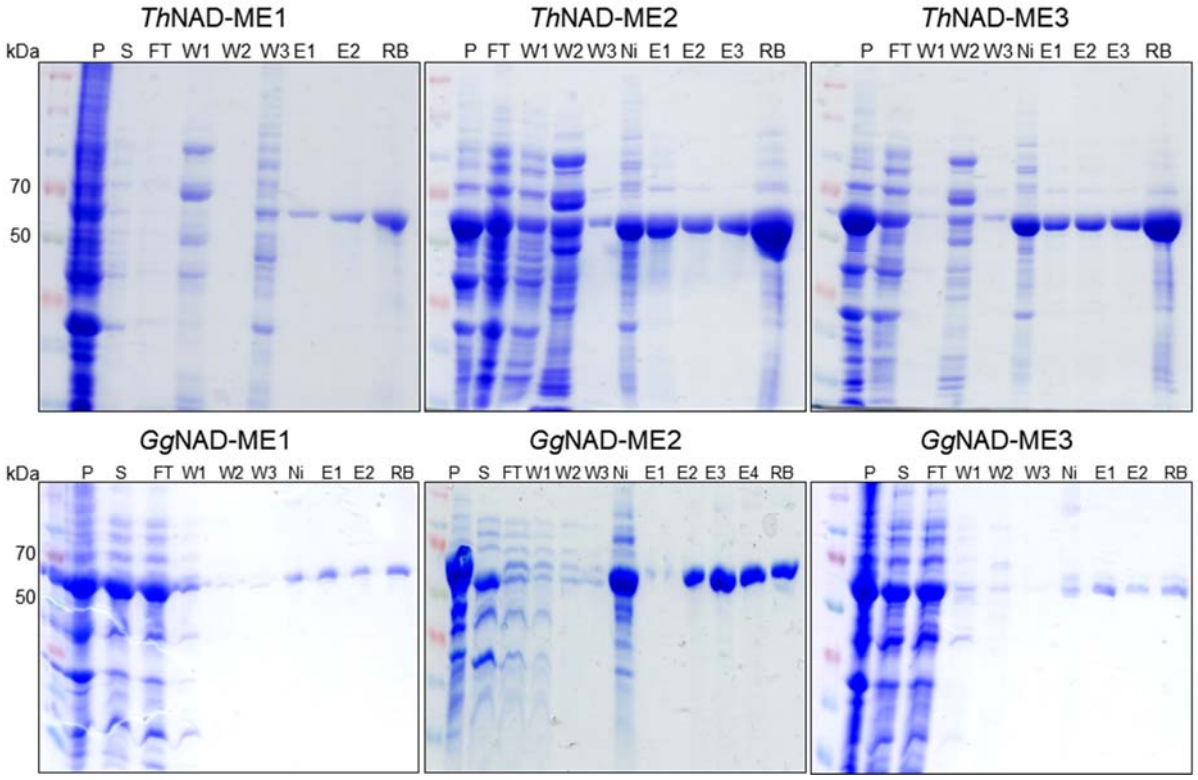
Worden AZ, Lee J, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett M V, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist E a, Panaud O, Pangilinan J, Paulsen I, Piegu B and Poliakov A. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268–272.

Xu Y, Bhargava G, Wu H, Loeber G and Tong L. (1999) Crystal structure of human mitochondrial $NAD(P)^+$-dependent malic enzyme: a new class of oxidative decarboxylases. *Structure* **7**, 877–889.

# Supplemental Figures



**Supplemental Figure 1.** Full maximum likelihood phylogenetic tree of 173 NAD-ME protein sequences. Values at nodes represent bootstrap support values (n=1000) in percent, branches with support ≤ 40% were collapsed. Green, α NAD-ME proteins; orange, β NAD-ME proteins.




**Supplemental Figure 2:** Representative SDS-PAGEs illustrating purification of the recombinant proteins to homogeneity. Expected molecular weights are ~64kDa of the Cleome mature protein sequence plus ~2.5 kDa of the HIS-Tag. From left to right: molecular weight marker, *E. coli* cell debris pellet after sonication (P), supernatant containing soluble protein after sonication (S), protein not binding to Ni-NTA agarose (flow through, FT), washing fractions (W1-3) with increasing imidazole concentration, protein bound to Ni-NTA agarose after elution (Ni), elution fractions (E1-4), concentrated protein in salt free buffer (rebuffered, RB).

# Supplemental Tables

**Supplemental Table 1.** List of primers used in this study

| purpose | name | sequence 5' to 3' |
|---|---|---|
| Cloning | ThME1_for | ATGGCGGTCGCTGGAGATAAG |
| Cloning | ThME1_rev | GTCATCTTTGTAGACCAAAGTCGG |
| Cloning | ThME2_into16bGibson | CATCATCATCATCATCATCATCACAGCAGCGGCCAT ATCGAAGGTCGTCATATGTGCATCGTCCACAAGCGA |
| Cloning | ThME2_into16bGibson | GCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAG CCGTCACTTCTCGTGAACCAGAG |
| Cloning | ThME2 for pET16 | GGCATATGTGCATCGTCCACAAGCGA |
| Cloning | ThME2 Cterm | CGGATCCTCACTTCTCGTGAACCAGAGGG |
| Cloning | ThME3 for | GCATATGTGCATTGTCCACAAGCGTG |
| Cloning | ThME3 rev | CCTCGAGTTATTTCTCGTGAACGAGAGGGC |
| Cloning | ThME3_into16bGibson | CATCATCATCATCATCATCATCACAGCAGCGGCCATA TCGAAGGTCGTCATATGTGCATTGTCCACAAGCGTG |
| Cloning | ThME3_into16bGibson | GCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAG CCGTTATTTCTCGTGAACGAGAGG |
| Cloning | GgNADME1fow | ATCCATGGCCCACCATTGTCCACAAGCGAA |
| Cloning | GgNADME1rev | ATGTCGACCTAAGCATTCTTGTAGACGAGAG |
| Cloning | GgME2 for pET | GCATATGTGCATCGTCCACAAGCGT |
| Cloning | GgME2 rev pET | CCTCGAGTCATTTTTTGTGAACCAAAGGG |
| Cloning | GgME3 rev | TCATTTCTCGTGAACCAGAGGGT |
| Cloning | GgME3 for pET16b | GCATATGTGTATTGTCCATAAGCGTGGC |
| Cloning | GgME3 rev pET16b | CGGATCCTCATTTCTCGTGAACCAGAGG |
| Cloning | GgME3_into16bGibson | CATCATCATCATCATCATCATCACAGCAGCGGCCATAT CGAAGGTCGTCATATGTGTATTGTCCATAAGCGTG |
| Cloning | GgME3_into16bGibson | GCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGC CGTCATTTCTCGTGAACCAG |
| qRT analysis | GgME1_for_qRT | ACTGCTTTTGATAGTGCACGAAG |
| qRT analysis | GgME1_rev_qRT | CTGGTGCTCGATTTCTTTAACC |
| qRT analysis | GgME2_for_qRT | GACACACTTGGAGGCCG |
| qRT analysis | GgME2_rev_qRT | CCAACTCCAGATAATCCCACA |
| qRT analysis | GgME2-2_for_qRT | GAGCAAAACTGGACTCAGCA |
| qRT analysis | GgME2-2_rev_qRT | ACCAACTCCAGATAATCCAAGA |
| qRT analysis | ThME1_for_qRT | TGCTTTTGATAGTGCACGAAGT |
| qRT analysis | ThME1_rev_qRT | GTCCCTGACGCTCAATTTCT |
| qRT analysis | ThME2_for_qRT | AGCTGAGATTTCAGGCCTATC |
| qRT analysis | ThME2_rev_qRT | CGGATTCTCGCATTGCCC |
| qRT analysis | ThME3_for_qRT | AGGTCAGGTTTCATGTCTTCGT |
| qRT analysis | ThME3_rev_qRT | TTCGAATCGGATCCTCGCAT |
| qRT analysis | Gg_Actin_for | GTCTTGACCTTGCAGGACG |
| qRT analysis | Gg_Actin_rev | AATCAAGGGCGACATACGAA |
| qRT analysis | Th_Actin_for | TCTACGAGGGTTATGCCCTT |
| qRT analysis | Th_Actin_rev | ATTTCACGTTCAGCAGTGGT |
| qRT analysis | GgME1_for_qRT_2 | TGGTTGATGCTCAGGGTCTT |
| qRT analysis | GgME1_rev_qRT_2 | CACTTGCCGAACCACTTCAG |
| qRT analysis | GgME2_for_qRT_2 | GTAGAAGTGGTGAAGAAGGTGAA |
| qRT analysis | GgME2_rev_qRT_2 | AGACATAGCAAAAATGGCAGGT |
| qRT analysis | GgME2_for_qRT_3 | GGTGAAGAAGGTGAAGCCGA |
| qRT analysis | GgME2_rev_qRT_3 | CAGTGCATTCGGCATTGGA |
| qRT analysis | GgME3_for_qRT_2 | TGCAATGTCGAATCCTACCATG |
| qRT analysis | GgME3_rev_qRT_2 | GTTACGATACGAGCACCAGATAAA |

# 3. Discussion

This thesis aimed at expanding knowledge on aspects of small molecule damage control systems in photosynthetic eukaryotes. Several studies were conducted on damage control systems specific for organisms performing oxygenic photosynthesis. Besides the discovery and description of these specific systems, ones that exists in similar fashion in animals or bacteria are described here. These systems found in all kingdoms of life offer a view on parts of metabolism that are likely to be universal.

## 3.1 A single step repair system for L-2-hydroxyglutarate

We discovered and described the single step repair system involving the repair enzyme L-2HGDH in *A. thaliana* (Hüdig *et al.*, 2015). L-2HGDH is a flavin-containing enzyme in plants and animals evolved to correct the metabolic error of mitochondrial malate dehydrogenase (mMDH). The main reaction of mMDH is the NAD-dependent interconversion of malate and OAA, but in a side reaction mMDH also reduces 2-ketoglutarate (2-KG) to L-2HG using NADH (Rzem *et al.*, 2004; Topçu *et al.*, 2004; Hüdig *et al.*, 2015). In animals as well as in *A. thaliana*, L-2HGDH returns the dead-end molecule L-2HG to central metabolism as 2-KG and donates the electrons to the mitochondrial electron transport chain (Hüdig *et al.*, 2015).

Despite the long evolutionary distance between plant and animal mitochondria and the probable metabolic changes that occurred along this way, both promiscuity of mMDH leading as well as its repair mechanism are conserved. The single human mMDH, MDH2 (Rzem *et al.*, 2007), as well as remarkably both mMDH enzymes in *A. thaliana*, mMDH1 and mMDH2, show promiscuity. The product of the reaction L-2HG is toxic for humans and loss-of-function mutations in the repair enzyme lead to severe damage in humans (Rzem *et al.*, 2004), while we found L-2HG accumulation to be non-toxic for *A. thaliana* in a wide range of tested conditions.

The continuous exhibition of the side reaction by both human and plant mMDH leads to the speculation that either mMDH is stuck on an evolutionary local maximum and cannot de-evolve this erroneous side reaction, or L-2HG, despite being toxic in the tested tissues in human, fulfills other non-toxic functions that are still to be discovered. While the maintenance of L-2HGDH in human mitochondria is of clear advantage, maintaining a single repair enzyme to return a dead-end molecule back to the metabolism seems to give a small but significant fitness advantage in plants. When

considering that all tested genomes of Viridiplantae possess an ortholog of the repair enzyme L-2HGDH, and the overall low but continuous expression of *AtL-2HGDH*, ancient universal need of the repair enzyme is likely.

Having a model organism not showing a severe response to primary a knock-out of the repair enzyme itself, like the *A. thaliana l-2hgdh1*, offers the possibility to track the influence of the damaged molecule and its downstream impact on metabolism and macromolecules. Hence, informing on intervention possibilities in organisms that show severe impairments upon loss of the repair enzyme. Further research on the metabolic influences of the damaged molecule or the evolutionary constrains of the participating enzymes in plants thus can inform about the mechanisms behind L-2HG toxicity in other eukaryotes.

## 3.2 Identification of a unique mitochondrial glycolate dehydrogenase participating in the photorespiratory pathway in *P. tricornutum*

Unicellular photosynthetic diatoms have next to photorespiration, as a repair mechanism for the toxic PG created by oxygenic photosynthesis, the possibility to prohibit damage by excretion of the likewise toxic intermediate of the photorespiratory pathway glycolate (Hellebust, 1965). The multi-step repair pathway constituting photorespiration was found to be different in land plants and cyanobacteria, involving different enzymes for glycolate to glyoxylate conversion. We found that consistent with the lack of $H_2O_2$ production in diatom peroxisomes (Winkler & Stabenau, 1995), glycolate is not oxidized in the peroxisomes in *P. tricornutum* (Schmitz *et al.*, 2017*a*). Instead, glycolate is converted by *Pt*GO1, a glycolate dehydrogenase that localizes to mitochondria and donates the reactions electrons to electron acceptors other than $O_2$. We showed that *Pt*GO1 is not evolutionary related to other known mitochondrial GlyDHs, e.g. from cyanobacteria, or peroxisomal GOX from plants. Additionally, we described that other proteins with low homology to glycolate oxidases, *Pt*PO2 as well as *Pt*DH1-3, do not participate in glycolate metabolism. *Pt*PO1 has unique enzymatic properties differing from plant GOX, comprising both L-lactate and glycolate conversion activity, but lacking activity on D-lactate. However, *Pt*PO1 seems to be restricted to a small group of heterokont algae that share orthologous protein sequences. *Pt*GO1 was likely acquired from an α-proteobacterium via horizontal gene transfer to the chlorophyte endosymbiont of the primary endosymbiosis in the ancestor

of the diatoms and heterokont algae. Thus, *Pt*PO1 represents a third way to metabolize photorespiratory glycolate (Schmitz *et al.*, 2017*a*).

Multi-step repair pathways are more likely to present different 'solutions' in different organisms as metabolic needs change. Diatoms are a group of organisms that share a secondary endosymbiotic event that replaced the chlorophyte endosymbiont with a rhodophyte endosymbiont. In this small group of organisms, the acquisition of a glycolate oxidase from the first endosymbiont enabled the common ancestor to perform the photorespiratory pathway without the generation of $H_2O_2$ in peroxisomes, a potential fitness advantage and reason for the loss of the GlyDH from the secondary endosymbiont. However, regardless which variant of the photorespiratory pathway is performed, the integration of this damage control pathway into the central metabolism constrains the flexibility of the pathway. This was also demonstrated by attempts to short-circuit photorespiration, where fine-tuning the changed enzymatic reactions in the new cellular compartment seemed to be the reason for differing biomass gains in different transgenic plant lines (Maier *et al.*, 2012). $H_2O_2$ production by the used GOX and its detoxification were identified as limiting factors. The use of a GlyDH like the one found in the diatom *P. tricornutum* could be a suitable substitute for the used chloroplast-only photorespiratory pathway. Hence, new variants of the photorespiratory pathway from other photosynthetic eukaryotes can open up possibilities for genetic engineering attempts in the field of photorespiration.

## 3.3 The scavenging system of methylglyoxal in *A. thaliana*

The free methylglyoxal scavenging system in plants involves two enzymes, GLXI and GLXII. They detoxify the GSH-scavenged methylglyoxal and regenerate the GSH pool, respectively. We described the physiological relevant isoforms of the multi-gene *At*GLXI and *At*GLXII families and identified their respective subcellular localization. *At*GLXI isoforms were heterologously expressed and their kinetic properties were characterized. We described a previously unknown metal co-factor dependent substrate preference and show that *At*GLXI isoforms respond differently to external stimuli by loss-of-function mutant analysis. This allowed us to propose for the first time a cellular model for the damage control system of the RCS methylglyoxal in *A. thaliana* (Schmitz *et al.*, 2017*b*). Eukaryotes performing oxygenic photosynthesis exhibit a second source of methylglyoxal generation through the TPI reaction in the chloroplast next to the TPI reaction in the cytosol as part of glycolysis. Interestingly, we found a

full detoxification system of GLXI and GLXII in both chloroplast and cytosol. Additionally, we found the second, GSH-releasing step to be also present in mitochondria, thus shifting the GSH bound to the intermediate from one compartment (cytosol and/or chloroplast) to mitochondria, where also the metabolization of the final produce D-lactate takes place (Maurino *et al.*, 2016; Schmitz *et al.*, 2017*b*).

## 3.4 $C_4$ photosynthesis, a multicellular steering system that evolved at least three times in Cleome

Another possibility that was evolved to deal with the oxygenase reaction of RubisCO is the suppression of this side reaction by concentrating $CO_2$ around RubisCO. There are multiple very different variants how this is achieved, e.g. carboxysomes in cyanobacteria, crassulacean acid metabolism (CAM) and $C_4$ photosynthesis. Contrasting to several well described $C_4$ photosynthesis variants depending on NADP-ME, is was unknown if a separate $C_4$ NAD-ME exists in plants exhibiting NAD-ME dependent $C_4$ photosynthesis. We found that in Cleome two different NAD-ME entities exists that differ in their subunit composition. Thus, we hypothesize that housekeeping and $C_4$ function, associated with different metabolic needs, are performed by different NAD-ME entities with specific kinetic properties. We show that a retained duplication of a *NAD-ME* gene aided the evolution of a new enzyme function, but the final $C_4$ NAD-ME is comprised of two different subunits (manuscript 2.5).

The carbon concentration around RubisCO in $C_4$ photosynthesis depends on the spatial separation of first and final $CO_2$ fixation, including transport of the fixed carbon and the crucial release of $CO_2$ by a decarboxylase. The recruitment of pre-existing enzymes into this pathway requires new possibilities for regulation, e.g. for the synchronization with photosynthetic activity. Additionally, changes to the metabolite pools that go along with the steering system will influence the reactions and enzyme kinetics. Here in the NAD-ME $C_4$ photosynthesis, we find also the challenge that new and housekeeping function of a heteromeric NAD-ME needed to be balanced with only one newly emerged subunit. A restriction that possibly lead to only three $C_4$ lineages in the Cleomaceae despite the shared WGD that led to the aiding gene duplication. How this regulation of NAD-ME is achieved in Cleome is part of future work.

## 3.5 Plant biochemical damage control systems – a review

The overall concept laid out in the introduction of this thesis was summarized with a focus on plant damage control systems and enzymatic repair, scavenging and steering systems (Hüdig *et al.*, 2018, manuscript 2.4). We described the emerging view of small molecule damage control as part of the metabolic network. We transferred and summarized repair and scavenging systems known from literature to recent new findings. We proposed steering systems as a category in molecule damage control, and offered a new perspective on known concepts such as $C_4$ photosynthesis as a tool for metabolite pool size control and therefore damage prevention (Hüdig *et al.*, 2018). The future work that is to be done in small molecule damage control is considerably large, as for the model plant *A. thaliana* there are close to 7000 uncharacterized enzymes and transporters (Niehaus *et al.*, 2015) and even if enzymes were described, oftentimes potential side reactions are not included. We describe algorithm-based approaches that were established to tackle this task computationally (Hüdig *et al.*, 2018). Future machine-based learning approaches will be helpful to give at least a good estimate when searching for promiscuity in enzymes and the relevant side substrates.

Nevertheless, we can only confirm such enzymatic reactions experimentally, as e.g. switches in metal cofactor and subsequent substrate changes as described for *At*GLXI cannot be predicted at the moment. The finding of new molecules and the discovery of their origin reaction will be sped up as technical methods for identification of molecules improve in throughput, precision and sensitivity. When combining elegantly the generated data sets of the -omics era, meaning that they are generated in a targeted way for machine-learning approaches, we will be able to complete the description of the cellular reaction network.

## 4. References


Apel K and Hirt H. (2004) Reactive oxygen species: metabolism, oxidative stress, and signal Transduction. *Annual Review of Plant Biology* **55**, 373–399.

Asada K. (2006) Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiology* **141**, 391–396.

Aubry S, Brown NJ and Hibberd JM. (2011) The role of proteins in $C_3$ plants prior to their recruitment into the $C_4$ pathway. *Journal of Experimental Botany* **62**, 3049–3059.

Bairoch A. (2000) The ENZYME database in 2000. *Nucleic Acids Research* **28**, 304–305.

Barker MS, Vogel H and Schranz ME. (2009) Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* **1**, 391–399.

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

Bayat S, Schranz ME, Roalson EH and Hall JC. (2018) Lessons from Cleomaceae, the sister of crucifers. *Trends in Plant Science* **23**, 808–821.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L and Schwede T. (2014) SWISS-MODEL : modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**, 252–258.

Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Mass J, Lercher MJ, Westhoff P, Hibberd JM and Weber APM. (2011) An mRNA blueprint for $C_4$ photosynthesis derived from comparative transcriptomics of closely related $C_3$ and $C_4$ species. *Plant Physiology* **155**, 142–156.

Bräutigam A and Gowik U. (2016) Photorespiration connects $C_3$ and $C_4$ photosynthesis. *Journal of Experimental Botany* **67**, 2953–2962.

von Caemmerer S and Furbank RT. (2016) Strategies for improving $C_4$ photosynthesis. *Current Opinion in Plant Biology* **31**, 125–134.

Chang GG and Tong L. (2003) Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry* **42**, 12721–12733.

Christin P-A, Boxall SF, Gregory R, Edwards EJ, Hartwell J and Osborne CP. (2013) Parallel recruitment of multiple genes into $C_4$ photosynthesis. *Genome Biology and Evolution* **5**, 2174–2187.

Christopher JT and Holtum J. (1996) Patterns of carbon partitioning in leaves of crassulacean acid metabolism species during deacidification. *Plant Physiology* **112**, 393–399.

Colinas M, Shaw H V., Loubéry S, Kaufmann M, Moulin M and Fitzpatrick TB. (2014) A pathway for repair of NAD(P)H in plants. *Journal of Biological Chemistry* **289**, 14692–14706.

Cook RM, Lindsay JG, Wilkins MB and Nimmo HG. (1995) Decarboxylation of malate in the crassulacean acid metabolism plant *Bryophyllum* (*Kalanchoë*) *fedtschenkoi*. *Plant Physiology* **109**, 1301–1307.

Cuypers A, Smeets K and Vangronsveld J.(2009) Heavy metal stress in plants. *In Plant Stress Biology,* 161–178. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Dever L V, Pearson M, Ireland RJ, Leegood RC and Lea PJ. (1998) The isolation and characterisation of a mutant of the $C_4$ plant *Amaranthus edulis* deficient in NAD-malic enzyme activity. *Planta* **206**, 649–656.

Drincovich MF, Lara M V, Andreo CS and Maurino VG. (2011) $C_4$ decarboxylases: different solutions for the same biochemical problem, the provision of $CO_2$ to Rubisco in the bundle sheath cells. *In $C_4$ Photosynthesis and Related $CO_2$ Concentrating Mechanisms, Vol. 32,* 277–300.

Edwards EJ, Osborne CP, Strömberg CAE, Smith SA and $C_4$ Grasses Consortium. (2010) The origins of $C_4$ grasslands: integrating evolutionary and ecosystem. *Science* **328**, 587–591.

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A and Hagemann M. (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proceedings of the National Academy of Sciences* **105**, 17199–17204.

Emanuelsson O, Nielsen H, Brunak S and Von Heijne G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005–1016.

Erb TJ, Frerichs-Revermann L, Fuchs G and Alber BE. (2010) The apparent malate synthase activity of *Rhodobacter sphaeroides* is due to two paralogous enzymes, (3*S*)-malyl-coenzyme A (CoA)/β-methylmalyl-CoA lyase and (3*S*)-malyl-CoA thioesterase. *Journal of Bacteriology* **192**, 1249–1258.

Esser C, Kuhn A, Groth G, Lercher MJ and Maurino VG. (2014) Plant and animal glycolate oxidases have a common eukaryotic ancestor and convergently duplicated to evolve long-chain 2-hydroxy acid oxidases. *Molecular Biology and Evolution* **31**, 1089–1101.

Farmer EE and Davoine C. (2007) Reactive electrophile species. *Current Opinion in Plant Biology* **10**, 380–386.

Feodorova TA, Voznesenskaya EV, Edwards GE and Roalson EH. (2010) Biogeographic patterns of diversification and the origins of $C_4$ in *Cleome* (Cleomaceae). *Systematic Botany* **35**, 811–826.

Foyer CH and Noctor G. (2005) Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. *Plant Cell* **17**, 1866–1875.

Frelin O, Huang L, Hasnain G, Jeffryes JG, Ziemak MJ, Rocca JR, Wang B, Rice J, Roje S, Yurgel SN, Gregory JF, Edison AS, Henry CS, de Crécy-Lagard V and Hanson AD. (2015) A directed-overflow and damage-control *N*-glycosidase in riboflavin biosynthesis. *Biochemical Journal* **466**, 137–145.

Furbank RT. (2011) Evolution of the $C_4$ photosynthetic mechanism: are there really three $C_4$ acid decarboxylation types? *Journal of Experimental Botany* **62**, 3103–3108.

Furbank RT and Hatch MD. (1987) Mechanism of $C_4$ photosynthesis. *Plant Physiology* **85**, 958–964.

Gardner HW. (1989) Oxygen radical chemistry of polyunsaturated fatty acids. *Free Radical Biology & Medicine* **7**, 65–86.

Gibson DG. (2011) Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology* **498**, 349–361.

Gill SS and Tuteja N. (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiology and Biochemistry* **48**, 909–930.

Grossman L, Lin CI and Ahn Y. (1998) Nucleotide excision repair in *Escherichia coli*. *DNA Damage and Repair* **1**, 11–27.

Grover SD and Wedding RT. (1982) Kinetic ramifications of the association-dissociation behavior of NAD malic enzyme. *Plant Physiology* **70**, 1169–1172.

Grover SD, Canellas PF and Wedding RT. (1981) Purification of NAD malic enzyme from potato and investigation of some physical and kinetic properties. *Archives of Biochemistry and Biophysics* **209**, 396–407.

Hall JC, Sytsma KJ and Iltis HH. (2002) Phylogeny of Capparaceae and Brassicacecae based on chloroplast sequence data. *American Journal of Botany* **89**, 1826–1842.

Hanson AD, Henry CS, Fiehn O and de Crécy-Lagard V. (2016) Metabolite damage and metabolite damage control in plants. *Annual Review of Plant Biology* **67**, 131–152.

Hanson DT. (2016) Breaking the rules of Rubisco catalysis. *Journal of Experimental Botany* **67**, 3180–3182.

Hatch MD. (1988) $C_4$ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta* **895**, 81–106.

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

Heckmann D, Schulze S, Denton AK, Gowik U, Westhoff P, Weber APM and Lercher MJ. (2013) Predicting $C_4$ photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**, 1579–1588.

Hellebust JA. (1965) Excretion of some organic compounds by marine phytoplankton. *Limnology and Oceanography* **10**, 192–206.

Hüdig M, Schmitz J, Engqvist MKM and Maurino VG. (2018) Biochemical control systems for small molecule damage in plants. *Plant Signaling & Behavior* **13**, 1–7.

Hüdig M, Tronconi MA, Poschmann G and Maurino VG. (in preparation) Evolution of NAD-malic enzyme from a TCA cycle-associated enzyme to a $C_4$ decarboxylase was aided by duplication of a β-subunit in the genus *Cleome*.

Hüdig M, Maier A, Scherrers I, Seidel L, Jansen EEW, Mettler-Altmann T, Engqvist MKM and Maurino VG. (2015) Plants possess a cyclic mitochondrial metabolic pathway similar to the mammalian metabolic repair mechanism Involving malate dehydrogenase and L-2-hydroxyglutarate dehydrogenase. *Plant and Cell Physiology* **56**, 1820–1830.

Jones DT, Taylor WR and Thornton JM. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**, 275–282.

Katoh K, Kuma KI, Toh H and Miyata T. (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**, 511–518.

Keller MA, Piedrafita G and Ralser M. (2015) The widespread role of non-enzymatic reactions in cellular metabolism. *Current Opinion in Biotechnology* **34**, 153–161.

Khersonsky O and S.Tawfik D. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry* **79**, 471–505.

Koteyeva NK, Voznesenskaya EV, Roalson EH and Edwards GE. (2011) Diversity in forms of $C_4$ in the genus *Cleome* (Cleomaceae). *Annals of Botany* **107**, 269–283.

Külahoglu C, Denton AK, Sommer M, Maß J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo E, Buell CR, Simon R, De Veylder L, Bräutigam A and Weber APM. (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae $C_3$ and $C_4$ Plant Species. *Plant Cell* **26**, 3243–3260.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549.

Kuraku S, Zmasek CM, Nishimura O and Katoh K. (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research* **41**, 22–28.

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

Landry LG, Chapple C and Last RL. (1995) Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiology* **109**, 1159–1166.

Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB, Piednoel M, Gundlach H, Van Bel M, Meyberg R, Vives C, Morata J, Symeonidi A, Hiss M, Muchero W, Kamisugi Y, Saleh O, Blanc G, Decker EL, van Gessel N, Grimwood J, Hayes RD, Graham SW, Gunter LE, McDaniel SF, Hoernstein SNW, Larsson A, Li FW, Perroud PF, Phillips J, Ranjan P, Rokshar DS, Rothfels CJ, Schneider L, Shu S, Stevenson DW, Thümmler F, Tillich M, Villarreal Aguilar JC, Widiez T, Wong GK-S, Wymore A, Zhang Y, Zimmer AD, Quatrano RS, Mayer KFX, Goodstein D, Casacuberta JM, Vandepoele K, Reski R, Cuming AC, Tuskan GA, Maumus F, Salse J, Schmutz J and Rensing SA. (2018) The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant Journal* **93**, 515–533.

Lauterbach M, Schmidt H, Billakurthi K, Hankeln T, Westhoff P, Gowik U and Kadereit G. (2017) *De novo* transcriptome assembly and comparison of $C_3$, $C_3$-$C_4$, and $C_4$ species of tribe Salsoleae (Chenopodiaceae). *Frontiers in Plant Science* **8**, 1–14.

Le DT, Tarrago L, Watanabe Y, Kaya A, Lee BC, Tran U, Nishiyama R, Fomenko DE, Gladyshev VN and Tran L-SP. (2013) Diversity of plant methionine sulfoxide reductases B and evolution of a form specific for free methionine sulfoxide. *PLoS ONE* **8**, 1–8.

Lerma-Ortiz C, Jeffryes JG, Cooper AJL, Niehaus TD, Thamm AMK, Frelin O, Aunins T, Fiehn O, de Crécy-Lagard V, Henry CS and Hanson AD. (2016) 'Nothing of chemistry disappears in biology': the Top 30 damage-prone endogenous metabolites. *Biochemical Society Transactions* **44**, 961–971.

Linster CL, Van Schaftingen E and Hanson AD. (2013) Metabolite damage and its repair or pre-emption. *Nature Chemical Biology* **9**, 72–80.

Long JJ, Wang J-L and Berry JO. (1994) Cloning and analysis of the $C_4$ photosynthetic NAD-dependent malic enzyme of amaranth mitochondria. *The Journal of Biological Chemistry* **269**, 2827–2833.

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

Maier A, Fahnenstich H, von Caemmerer S, Engqvist MKM, Weber APM, Flügge U-I and Maurino VG. (2012) Transgenic introduction of a glycolate oxidative cycle into *A. thaliana* chloroplasts leads to growth improvement. *Frontiers in Plant Science* **3**, 1–12.

Mano J. (2012) Reactive carbonyl species: their production from lipid peroxides, action in environmental stress, and the detoxification mechanism. *Plant Physiology and Biochemistry* **59**, 90–97.

Marbaix AY, Noël G, Detroux AM, Vertommen D, Van Schaftingen E and Linster CL. (2011) Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide. *Journal of Biological Chemistry* **286**, 41246–41252.

Mary J, Vougier S, Picot CR, Perichon M, Petropoulos I and Friguet B. (2004) Enzymatic reactions involved in the repair of oxidized proteins. *Experimental Gerontology* **39**, 1117–1123.

Matsuoka M, Furbank RT, Fukayama H and Miyao M. (2001) Molecular engineering of $C_4$ photosynthesis. *Annual Review Of Plant Physiology and Plant Molecular Biology* **52**, 297–314.

Maurino VG and Peterhänsel C. (2010) Photorespiration: current status and approaches for metabolic engineering. *Current Opinion in Plant Biology* **13**, 249–256.

Michal G. (2014) *Part 1 Metabolic Pathways 4th edition*. F. Hoffmann-La Roche Ltd. Downloaded from http://biochemical-pathways.com/#/map/1 on 8 October 2018.

Monson RK. (2003) Gene duplication, neofunctionalization, and the evolution of $C_4$ photosynthesis. *International Journal of Plant Science* **164**, 43–54.

Moreno-Villena JJ, Dunning LT, Osborne CP and Christin PA. (2018) Highly expressed genes are preferentially co-opted for $C_4$ photosynthesis. *Molecular Biology and Evolution* **35**, 94–106.

Murata T, Ohsugi R, Matsuoka M and Nakamoto H. (1989) Purification and characterization of NAD malic enzyme from leaves of *Eleusine coracana* and *Panicum dichotomiflorum*. *Plant Physiology* **89**, 316–324.

Murphy MP. (2009) How mitochondria produce reactive oxygen species. *Biochemical Journal* **417**, 1–13.

Nakamura Y, Kanakagiri S, Van K, He W and Spalding MH. (2005) Disruption of the glycolate dehydrogenase gene in the high-$CO_2$-requiring mutant HCR89 of *Chlamydomonas reinhardtii*. *Canadian Journal of Botany* **83**, 820–833.

Niehaus TD, Thamm AM, de Crécy-Lagard V and Hanson AD. (2015) Proteins of unknown biochemical function: a persistent problem and a roadmap to help overcome it. *Plant Physiology* **169**, 1436–1442.

Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, Vosolsobě S, Rombauts S, Wilhelmsson PKI, Janitza P, Kern R, Heyl A, Rümpler F, Calderón Villalobos LIA, Clay JM, Skokan R, Toyoda A, Suzuki Y, Kagoshima H, Schijlen E, Tajeshwar N, Catarino B, Hetherington AJ, Saltykova A, Bonnot C, Breuninger H, Symeonidi A, Radhakrishnan GV, Van Nieuwerburgh F, Deforce D, Chang C, Karol KG, Hedrich R, Ulvskov P, Glöckner G, Delwiche CF, Petrášek J, Van de Peer Y, Friml J, Beilby M, Dolan L, Kohara Y, Sugano S, Fujiyama A, Delaux PM, Quint M, Theißen G, Hagemann M, Harholt J, Dunand C, Zachgo S, Langdale J, Maumus F, Van Der Straeten D, Gould SB and Rensing SA. (2018) The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448–464.

Ohsugi R and Murata T. (1980) Leaf anatomy, post-illumination $CO_2$ burst and NAD-malic enzyme activity of *Panicum dichotomiflorum*. *Plant and Cell Physiology* **21**, 1329–1333.

Oppenheimer NJ and Kaplan NO. (1974) Glyceraldehyde-3-phosphate dehydrogenase catalyzed hydration of the 5-6 double bond of reduced *β*-nicotinamide adenine dinucleotide (*β*NADH). Formation of *β*-6-hydroxy-1, 4, 5, 6-tetrahydronicotinamide adenine dinucleotide. *Biochemistry* **13**, 4685–4694.

Peterhänsel C and Maurino VG. (2011) Photorespiration redesigned. *Plant Physiology* **155**, 49–55.

Piedrafita G, Keller MA and Ralser M. (2015) The impact of non-enzymatic reactions and enzyme promiscuity on cellular metabolism during (oxidative) stress conditions. *Biomolecules* **5**, 2101–2122.

Prabhakar P, Laboy JI, Wang J, Budker T, Din ZZ, Chobanian M and Fahien LA. (1998) Effect of NADH-X on cytosolic glycerol-3-phosphate dehydrogenase. *Archives of Biochemistry and Biophysics* **360**, 195–205.

Rangan P, Furtado A and Henry RJ. (2016) New evidence for grain specific $C_4$ photosynthesis in wheat. *Scientific Reports* **6**, 31721.

Reaves ML, Young BD, Hosios AM, Xu Y-F and Rabinowitz JD. (2013) Pyrimidine homeostasis is accomplished by directed overflow metabolism. *Nature* **500**, 237–241.

Richard JP. (1993) Mechanism for the formation of methylglyoxal from triosephosphates. *Biochemical Society Transactions* **21**, 549–553.

Rzem R, Veiga-da-Cunha M, Noël G, Goffette S, Nassogne M-C, Tabarki B, Schöller C, Marquardt T, Vikkula M and Van Schaftingen E. (2004) A gene encoding a putative FAD-dependent L-2-hydroxyglutarate dehydrogenase is mutated in L-2-hydroxyglutaric aciduria. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16849–16854.

Rzem R, Vincent MF, Van Schaftingen E and Veiga-da-Cunha M. (2007) L-2-hydroxyglutaric aciduria, a disorder of metabolite repair. *Journal of Inherited Metabolic Disease* **30**, 681–689.

Sage RF. (2017) A portrait of the $C_4$ photosynthetic family on the $50^{th}$ anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of Experimental Botany* **68**, e11–e28.

Sage RF, Christin P-A and Edwards EJ. (2011) The $C_4$ plant lineages of planet Earth. *Journal of Experimental Botany* **62**, 3155–3169.

Sage RF, Sage TL and Kocacinar F. (2012) Photorespiration and the evolution of $C_4$ photosynthesis. *Annual Review of Plant Biology* **63**, 19–47.

Saigo M, Bologna FP, Maurino VG, Detarsio E, Andreo CS and Drincovich MF. (2004) Maize recombinant non-$C_4$ NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Molecular Biology* **55**, 97–107.

Sandalio LM, Romero-puertas C and Luis AR. (2013) Role of peroxisomes as a source of reactive oxygen species (ROS) signaling molecules. *In Peroxisomes and Their Key Role in Cellular Signaling and Metabolism, Vol. 69,* 231–255. Springer Science+Business Media, Dordrecht.

Satoh K. (1995) The high non-enzymatic conjugation rates of some glutathione *S*-transferase (GST) substrates at high glutathione concentrations. *Carcinogenesis* **16**, 869–874.

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

Van Schaftingen E, Rzem R, Marbaix A, Collard F, Veiga-Da-Cunha M and Linster CL. (2013) Metabolite proofreading, a neglected aspect of intermediary metabolism. *Journal of Inherited Metabolic Disease* **36**, 427–434.

Schmitz J, Srikanth NV, Hüdig M, Poschmann G, Lercher MJ and Maurino VG. (2017a) The ancestors of diatoms evolved a unique mitochondrial dehydrogenase to oxidize photorespiratory glycolate. *Photosynthesis Research* **132**, 183–196.

Schmitz J, Dittmar IC, Brockmann JD, Schmidt M, Hüdig M, Rossoni AW and Maurino VG. (2017b) Defense against reactive carbonyl species involves at least three subcellular compartments where individual components of the system respond to cellular sugar status. *Plant Cell* **29**, 3234–3254.

Schranz ME and Mitchell-Olds T. (2006) Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* **18**, 1152–1165.

Schwander T, Schada von Borzyskowski L, Burgener S, Cortina NS and Erb TJ. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354**, 900–904.

Stadtman ER. (1993) Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. *Annual Review of Biochemistry* **62**, 797–821.

Sun J, Jeffryes JG, Henry CS, Bruner SD and Hanson AD. (2017) Metabolite damage and repair in metabolic engineering design. *Metabolic Engineering* **44**, 150–159.

Tachiev G, Roth JA and Bowers AR. (2000) Kinetics of hydrogen peroxide decomposition with complexed and ‘free’ iron catalysts. *International Journal of Chemical Kinetics* **32**, 24–35.

Tausta SL, Miller Coyle H, Rothermel B, Stiefel V and Nelson T. (2002) Maize $C_4$ and non-$C_4$ NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Molecular Biology* **50**, 635–652.

Team R Core. (2017) R: A language and environment for statistical computing.

Thornalley PJ. (1990) The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochemical Journal* **269**, 1–11.

Timm S, Nunes Nesi A, Parnik T, Morgenthal K, Wienkoop S, Keerberg O, Weckwerth W, Kleczkowski LA, Fernie AR and Bauwe H. (2008) A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in *Arabidopsis*. *Plant Cell* **20**, 2848–2859.

Topçu M, Jobard F, Halliez S, Coskun T, Yalçinkayal C, Gerceker FO, Wanders RJA, Prud'homme JF, Lathrop M, Özguc M and Fischer J. (2004) L-2-hydroxyglutaric aciduria: identification of a mutant gene C14orf160, localized on chromosome 14q22.1. *Human Molecular Genetics* **13**, 2803–2811.

Torres MA. (2010) ROS in biotic interactions. *Physiologia Plantarum* **138**, 414–429.

Tronconi MA, Andreo CS and Drincovich MF. (2018) Chimeric structure of plant malic enzyme family: Different evolutionary scenarios for NAD- and NADP-dependent isoforms. *Frontiers in Plant Science* **9**, 1–15.

Tronconi MA, Maurino VG, Andreo CS and Drincovich MF. (2010a) Three different and tissue-specific NAD-malic enzymes generated by alternative subunit association in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **285**, 11870–11879.

Tronconi MA, Gerrard Wheeler MC, Maurino VG, Drincovich MF and Andreo CS. (2010b) NAD-malic enzymes of *Arabidopsis thaliana* display distinct kinetic mechanisms that support differences in physiological control. *The Biochemical Journal* **430**, 295–303.

Tronconi MA, Wheeler MCG, Martinatto A, Zubimendi JP, Andreo CS and Drincovich MF. (2015) Allosteric substrate inhibition of Arabidopsis NAD-dependent malic enzyme 1 is released by fumarate. *Phytochemistry* **111**, 37–47.

Tronconi MA, Fahnenstich H, Gerrard Wheeler MC, Andreo CS, Flügge U-I, Drincovich MF and Maurino VG. (2008) Arabidopsis NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact on nocturnal metabolism. *Plant Physiology* **146**, 1540–1552.

Vieira Dos Santos C, Cuiné S, Rouhier N and Rey P. (2005) The Arabidopsis plastidic methionine sulfoxide reductase B proteins. Sequence and activity characteristics, comparison of the expression with plastidic methionine sulfoxide reductase A, and induction by photooxidative stress. *Plant Physiology* **138**, 909–922.

Vistoli G, De Maddis D, Cipak A, Zarkovic N, Carini M and Aldini G. (2013) Advanced glycoxidation and lipoxidation end products (AGEs and ALEs): an overview of their mechanisms of formation. *Free Radical Research* **47**, 3–27.

Wang Y, Long SP and Zhu X-G. (2014a) Elements required for an efficient NADP-malic enzyme type $C_4$ photosynthesis. *Plant Physiology* **164**, 2231–2246.

Wang Y, Bräutigam A, Weber APM and Zhu X-G. (2014b) Three distinct biochemical subtypes of $C_4$ photosynthesis? A modelling analysis. *Journal of Experimental Botany* **65**, 3567–3578.

Warner DA, Ku MSB and Edwards GE. (1987) Photosynthesis, leaf anatomy, and cellular constituents in the polyploid $C_4$ grass *Panicum virgatum*. *Plant Physiology* **84**, 461–466.

Welchen E, García L, Schmitz J, Fuchs P, Wagner S, Wienstroer J, Schertl P, Braun H-P, Schwarzländer M, Gonzalez DH and Maurino VG. (2016) D-lactate dehydrogenase links methylglyoxal degradation and electron transport through cytochrome C. *Plant Physiology* **172**, 901–912.

Winkler U and Stabenau H. (1995) Isolation and characterization of peroxisomes from diatoms. *Planta* **195**, 403–407.

Worden AZ, Lee J, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett M V, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist E a, Panaud O, Pangilinan J, Paulsen I, Piegu B and Poliakov A. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268–272.

Xu Y, Bhargava G, Wu H, Loeber G and Tong L. (1999) Crystal structure of human mitochondrial $NAD(P)^+$-dependent malic enzyme: a new class of oxidative decarboxylases. *Structure* **7**, 877–889.

Yoshida A and Dave V. (1975) Inhibition of NADP-dependent dehydrogenases by modified products of NADPH. *Archives of Biochemistry and Biophysics* **169**, 298–303.

## 5. Acknowledgements

First and foremost, I would like to thank my supervisor PD Dr. Veronica Maurino. She made me feel welcome from the beginning, and created a unique atmosphere in the working group. Throughout the years, her trust in me made me grow as a person and scientist and gave me space to develop further.

I thank Prof. Dr. Martin Lercher for kindly agreeing to read and grade my thesis. The intense discussions throughout the years were exceptionally helpful and always offered new perspectives on the discussed matter. The openness and friendly atmosphere is also mirrored in his group: fostering more than one scientific collaboration, successfully connecting theoreticians and wet lab scientists.

Furthermore, I would like to thank Prof. Dr. Peter Westhoff for being my mentor. His interest in my topic, and the crisp discussions during meetings, helped to advance my projects.

I also would like to acknowledge being given the opportunity to work with and in such a fantastic group of people, the Maurino lab. Jessi, you are such a keen observer with an incredible knowledge and an eye for details when it comes to science, but furthermore you now also have become a friend. Anastasiia, you started your PhD journey shortly after me, and having a fighter next to you in a similar situation makes everything a bit easier – thank you so much. Alex and Alex, who introduced me to parts of the projects. Colleagues that came and went: Pia, Yuanyuan, Sabine – all of you contributed to the lively atmosphere that I value so much and let us not forget you moving the lab with us (what crazy times!). Also, I would like to thank my students Elena, Caren and Fatiha who furthered my other projects and helped me in becoming a better supervisor.

Further collaborators were Gereon, performing the MS runs, and Marcos for his collaboration on the Cleome project. I received general support from various people in the Weber lab (chocolate, drinks, and discussions) and I still appreciate the friendliness when sharing equipment.

Of course, no thesis would be complete without the acknowledgment of funding: I therefore would like to thank CEPLAS for the funds for the first three years of my work, but even more for creating a platform to meet exceptional young scientists from various backgrounds and with strong and interesting perspectives on life, science and

everything. The diversity that I encountered enriched my personal life and helped spark new research ideas. Sometimes with the help of a glass of wine (WW group I love you).

In addition to the support of science ideas, funding of my work and lab supplies, I also needed tons of moral support. I therefore would like to thank old friends of mine, Stephie, Marc, Natalie, you got my back no matter what obstacles appeared in my way. Stephan R. for helping me by very timely proofreads. The new friends I made in Düsseldorf, Esther, Katharina, Steffi, you understood exactly what it meant when yet another experiment went down the drain. You hugged me tightly until everything stopped looking so bad and at other times celebrated the small and the big victories with me.

And last, but not least, I would like to thank my family for their everlasting unconditional support. Your straightforwardness in all things, and which you also taught me, made me the person I am today.