Repertoire of malic enzymes in yeast and fungi: insight into their evolutionary functional and structural significance

Tayvich Vorapreeda,1 Chinea Thammarongtham,1 Supapon Cheevadhanarak2,3 and Kobkul Laoteng4

1Biochemical Engineering and Pilot Plant Research and Development Unit, National Center for Genetic Engineering and Biotechnology, King Mongkut's University of Technology Thonburi, Bangkhunthien, Bangkok 10150, Thailand
2School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkhunthien, Bangkok 10150, Thailand
3Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhunthien, Bangkok 10150, Thailand
4Bioresources Technology Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Thailand Science Park, Pathum Thani 12120, Thailand

Malic enzyme (ME) is one of the important enzymes for furnishing the cofactor NAD(P)H for the biosynthesis of fatty acids and sterols. Due to the existence of multiple ME isoforms in a range of oleaginous microbes, a molecular basis for the evolutionary relationships amongst the enzymes in oleaginous fungi was investigated using sequence analysis and structural modelling. Evolutionary distance and structural characteristics were used to discriminate the MEs of yeasts and fungi into several groups. Interestingly, the NADP+−dependent MEs of Mucoromycotina had an unusual insertion region (FLxxPG) that was not found in other fungi. However, the subcellular compartment of the Mucoromycotina enzyme could not be clearly identified by an analysis of signal peptide sequences. A constructed structural model of the ME of Mucor circinelloides suggested that the insertion region is located at the N-terminus of the enzyme (aa 159–163). In addition, it is presumably part of the dimer interface region of the enzyme, which might provide a continuously positively charged pocket for the efficient binding of negatively charged effector molecules. The discovery of the unique structure of the Mucoromycotina ME suggests the insertion region could be involved in particular kinetics of this enzyme, which may indicate its involvement in the lipogenesis of industrially important oleaginous microbes.

INTRODUCTION

Malic enzyme (ME) is essential for catalysing the reversible oxidative decarboxylation of L-malate in conjunction with the reduction of NAD(P)+ to NAD(P)H. The catalytic function of this enzyme requires a divalent metal cation (Mg2+ or Mn2+). MEs are involved in a wide range of metabolic reactions in living cells, and thus are classified into three categories according to their cofactor preference and substrate specificity. The NADP+−dependent enzyme (NADP+−ME, EC 1.1.1.40), which preferentially utilizes NADP+, has the ability to decarboxylate oxaloacetate (OAA), and has been found in cytoplasm, mitochondria and chloroplast stroma. This enzyme is distributed widely in animals (Xu et al., 1999) and plants (Rothermel & Nelson, 1989). Another set of ME (NAD+−ME, EC 1.1.1.39) prefers NAD+ and cannot decarboxylate OAA. This group is localized to the mitochondrial matrix and has been found in many plant species (Winning et al., 1994). The last category of ME (NAD+−ME, EC 1.1.1.38), found across organisms including both prokaryotes and eukaryotes, uses preferentially NAD+ as an essential cofactor and is able to decarboxylate OAA.

In terms of their crystal structure, MEs are homotetrameric proteins with a double-dimer structure. It has been shown that the dimer interface facilitates stronger interaction than the tetramer interface (Hsieh et al., 2009). Each monomer structure is composed of four domains (A, B, C and D). Domain A is associated with dimer formation, and domain
D contains the C-terminal extension residues involved in interactions with another dimer.Domains B and C serve primarily as catalytic sites. Although the various MEs share similar tertiary and quaternary structures and a common catalytic mechanism, some different characteristics have been found, particularly in terms of catalytic regulation and redox cofactor specificity. For example, a human NAD(P)^+‐dependent ME has dual specificity for the dinucleotide cofactor NAD^+ and NADP^+ in the binding of malate, and is regulated by fumarate and ATP, whereas another isoform is specific for NADP^+.

Additionally, distinct roles of tetramer organization have been reported for the NADP^+‐ and NAD^+‐dependent ME isoforms (Yang et al., 2002; Tao et al., 2003; Hsieh et al., 2009).

Although the structure, substrate specificity, cofactor preference and compartmentalization of several MEs have been elucidated, the physiological role of the enzyme in individual organisms remains uncertain. In oleaginous micro-organisms, there is evidence demonstrating that the supply of NADPH through the catabolic reaction of ME is a favourable and essential route for lipid accumulation (Wynn & Ratledge, 1997; Wynn et al., 1999; Wang et al., 2011). The activity of ME correlates positively with the rate of fatty acid synthesis in many oleaginous microbes. In Aspergillus nidulans, cellular lipid accumulation was reduced by half as a result of a lack of ME activity (Wynn & Ratledge, 1997). Moreover, the lipid content of the Muco circinelloides increased from 12% to 30% of biomass when ME activity was increased two- to threefold (Zhang et al., 2007). In addition to the supply of the C2-metabolite acetyl-CoA, NADPH metabolism has been thought to be a target not only for controlling the production of microbial lipids, but also for manipulating metabolic flux for the biosynthesis of other hydrocarbon metabolites of commercial interest (Zelle et al., 2008, 2011). Furthermore, certain oleaginous fungi contain several isoforms of MEs, e.g. Muco circinelloides has six isoforms (Song et al., 2001) and Mortierella alpina has at least seven isoforms (Zhang & Ratledge, 2008). A homology search of the genome data of Muco circinelloides CBS277.49 showed that it contains five genes coding for MEs (Vongsangnak et al., 2012). The necessity of the existence of multiple isozymes in the cells and the contributions of each enzyme isoform to cellular processes mostly remain elusive. The availability of a greater range of sequence and structural data than generated so far would provide a better understanding of the structural and functional significance of this enzyme.

In this study, using a bioinformatics approach, we investigated fungal MEs based on phylogenetic relationship and subcellular localization, thereby inferring a divergence of the enzymic groups. In addition, the ME of Mucor circinelloides, having an unusual feature, was analysed intensively through structural modelling of the representative enzyme of Muco circinelloides. This study provides an insight into the biological evolution of MEs as well as their functional role in oleaginous micro-organisms.

### METHODS

**Retrieval of ME sequences.** The sequences of the MEs of Muco circinelloides and Rhizopus oryzae were retrieved from the genome databases of Muco circinelloides CBS277.49, V2.0 (http://genome.jgi-psf.org/Muc2/Muc2.download.html) and R. oryzae RA99-880 (http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/MultiHome.html) (Machida et al., 2005), respectively. The amino acid sequences of the MEs of other yeasts and fungi were retrieved from the National Center for Biotechnology Information database (http://ncbi.nlm.nih.gov).

**Phylogenetic tree analysis.** Multiple sequence alignments of the amino acid sequences were utilized to calculate the genetic distance using the MUSCLE program with the default settings (Edgar, 2004) in the MEGA5 software program (http://www.megasoftware.net/) (Tamura et al., 2011). An evolutionary tree was reconstructed using the maximum-likelihood method in conjunction with the Whelan and Goldman substitution model using a discrete Gamma distribution (+ G) with five rate categories under the assumption that a certain fraction of sites are evolutionarily invariable (+I) and these sites were selected as the best-fit substitution model of sequence evolution. The confidence intervals of the phylogenetic trees were tested using the bootstrap statistical method (Felsenstein, 1985) with 1000 resampling iterations.

**Prediction of subcellular localization of proteins.** Protein localization was analysed using the support vector machine system implemented in the program CELLO version 2.5 (Yu et al., 2006). The standard statistical values were used for the CELLO algorithm training by multiple feature vectors based on multiple n-peptide composition, which provides the accuracy of predictive results. If protein localization could not be identified by CELLO with a highly significantly predictive performance, TargetP 1.1 (Emmanuelsson et al., 2000) with the default parameter settings of the program was then used to analyse the potential N-terminal target sites for protein transport.

**Identification of conserved motifs.** Conserved sequence motifs were extracted from an alignment of amino acid sequences using the multi-sequence alignment program available on the MAFFT software program (http://mafft.cbrc.jp/alignment/software/) (Katoh & Toh, 2008) under the FFT-NS-i algorithm (iterative refinement method). The process was automatically repeated for a maximum of 1000 iterations or until no improvements were made in scoring alignment.

**Homology modelling of ME structure.** The molecular structures of the MEs of Muco circinelloides were modelled using a restraint-based modelling method implemented in the program MODELLER (Eswar et al., 2006). The local dynamic programming algorithm (Altschul et al., 1997) integrated in MODELLER was used for searching the modelling templates. Several models were generated and energy minimization was performed using the CHARMM-22 molecular mechanics force field (MacKerell et al., 1998). The final model was selected based on the lowest value of the MODELLER objective function and was considered the best one. The structural models were visualized using the Chimera package version 1.6.1 (http://www.cgl.ucsf.edu/chimera/) (Pettersen et al., 2004).

### RESULTS AND DISCUSSION

**Classification of yeast and fungal MEs**

Based on the maximum-likelihood method, we reconstructed a phylogenetic tree for the ME sequences of yeasts
and fungi, which included 68 sequences of 45 species as shown in Fig. 1. The consensus tree, which was measured by bootstrap analysis of 1000 replicates of random sequences, presented the yeast and fungal ME sequences as grouped together and clearly separated from the prokaryotic MEs (outgroups 1 and 2 of Escherichia coli and Synnechocystis sp., respectively). The sequences of yeasts and fungi were classified into 11 clusters: cluster A (subphylum Pezizomycotina, class Onygenales), cluster B (subphylum Pezizomycotina, class Eurotiales), cluster C (subphylum Pezizomycotina, class Pleosporales), cluster D (subphylum Mucoromycotina), cluster E (subphylum Pezizomycotina, class Sordariales), cluster F (subphylum Neocallimastigomycotina), cluster G (subphylum Mucoromycotina), cluster H (subphylum Agaricomycotina), cluster I (subphylum Pezizomycotina containing mixed classes), cluster J (subphylum Taphrinomycotina) and cluster K (subphylum Saccharomycotina). Clusters C, E, F and H each had only one member: EDU41282 (Pyrenophora tritici-repentis Pt-1C-BFP), CAP73752 (Podospora anserina Smat+), P78715 (Neocallimastix frontalis) and AAW45546 (Cryptococcus neoformans var. neoformans JEC21), respectively.

It has been reported that NAD\(^+\) and NADP\(^+\) play pivotal roles as reductants in concert with the catalytic function of MEs. Based on the sequence analysis of some MEs with known specific cofactors in this work, the yeast and fungal MEs were classified according to their reductant utilization. Considering the phylogenetic tree for subphylum Pezizomycotina, the MEs of cluster A, B, C and E were presumed to be NADP\(^+\)-dependent enzymes (EC 1.1.1.40), whereas the cluster I sequences were classified as putative NAD\(^+\)-dependent enzymes (EC 1.1.1.38), as shown in Fig. 1 and Supplementary Data S1 (available in Microbiology Online). In addition, NADP\(^+\)-dependent enzymes were also distributed in cluster F (van der Giezen et al., 1997), and the MEs belonging to the clusters H and J were NAD\(^+\)-dependent. Based on the genetic distance, the MEs of Mucoromycotina could be classified into two groups (clusters D and G), which is consistent with a very recent study of Mu. circinelloides (Vongsanong et al., 2012). However, it was unclear which cofactor participated in the catalytic function of the Mucoromycotina enzymes. In some cases, the MEs from the same fungal strains, such as Aspergillus oryzae RIB40 and Mucoromycotina fungi, were categorized in different clusters. Previously, an experimental study of Aspergillus spp. indicated the presence of multiple enzymes using different cofactors: MaeA (NADP\(^+\)-dependent malate dehydrogenase) and MaeB (NAD\(^+\)-dependent malate dehydrogenase) (Filippi et al., 2009). It is likely that the cofactor utilization of the MEs is not strictly dependent on either fungal group or strain. The largest cluster studied (cluster K) was the Saccharomycotina MEs; most of them contained only one NAD(P)\(^+\)-dependent ME-encoding gene (Boles et al., 1998; Tang et al., 2010) except for the ME of Vanderwaltozyma polyspora. It is remarkable that both non-oleaginous yeasts (i.e. Saccharomyces cerevisiae and Candida albicans) and oleaginous yeasts (i.e. Lipomyces starkeyi and Yarrowia lipolytica) were grouped together into cluster K. There have been reports of redundancy in reductant utilization by the S. cerevisiae and L. starkeyi MEs enabling them to employ either NADH or NADPH, but NADH is the preferred coenzyme (Boles et al., 1998; Tang et al., 2010).

As a result of the signal sequence analysis using Cello and TargetP (Supplementary Data S2), the MEs could be classified based on their subcellular localizations, including the cytosol and mitochondria. The NADP\(^+\)-dependent enzymes of clusters A, B, C and E were putatively localized in the mitochondria, whereas the subcellular compartmentalization of the NAD\(^+\)-dependent enzymes of clusters H and J was presumably in the cytosol. In addition, a mitochondrial location was found for the evolutionarily closely related enzyme of cluster I. The cluster F enzyme was predicted to be located in the mitochondrion, but it has been reported that this enzyme is located in a special organelle called the ‘hydrogenosome’. There is a close relationship between the mitochondria and hydrogenosome in terms of the protein import machineries, as reported previously (van der Giezen et al., 1997). In contrast with the other clusters, the largest cluster, cluster K (Saccharomycotina), included enzymes located in both the mitochondria and cytosol. The cluster K enzymes were mostly putative mitochondrial proteins (Hong et al., 2011) except for ADI76993.1 (L. starkeyi) and ABB55270 (Pichia jadinii), which were predicted to be cytosolic enzymes. This finding coincides with the previous report mentioning that the typical mitochondrial targeting sequence was absent in L. starkeyi ME (Tang et al., 2010). For the Mucoromycotina, the compartments could not be identified clearly using signal sequence analysis due to the aforementioned non-clarified cofactor utilization.

**Identification of conserved motifs in the MEs**

Using MAFFT software under the FFT-NS-i algorithm, a multiple alignment of the amino acid sequences of MEs (Supplementary Data S3) was able to identify the significantly conserved motifs, as shown in Fig. 2. Three conserved motifs, including two dinucleotide binding signature motifs and a divalent metal ion binding motif, were present in all clusters. The so-called divalent metal ion binding motif, (F/Y)ED-......-FD, was found between the two dinucleotide binding signature motifs, similar to the previous report describing its involvement in the binding of divalent metal ions, such as Mn\(^{2+}\) and Mg\(^{2+}\) (Zhang et al., 2007). The ‘dinucleotide binding signature’, named for the domain containing two motifs of GxG/....-GxG/...., was found in all clusters. For the second motif, all clusters except cluster D contained the GxGXXG pattern, whereas a GxGXXA motif was observed in...
cluster D. It has been postulated that the GxGxxG signature motif is involved generally in NAD$^+$ specificity, whereas the GxGxxA motif denotes specificity for the NADP$^+$ reductant (Hurley et al., 1996). However, the MEs appear to deviate from this rule. The GxGxxG motif is found in plant, yeast and bacterial MEs, whereas animal MEs contain the GxGxxA motif, irrespective of their cofactor specificity (Xu et al., 1999). In addition, the mutated enzyme of maize generated by site-directed mutagenesis of alanine to glycine (GxGxxA to GxGxxG) indicated that there was no major change in the structure or cofactor preference of the maize NADP$^+$-ME (Detarsio et al., 2003). In Mucoromycotina MEs, the cofactor specificity or preference could hardly be distinguished between clusters D and G based on the pattern of dinucleotide binding signature motifs because there was no relationship between the dinucleotide binding signature motifs and cofactor specificity and preference (Supplementary Data S1), which is similar to the previous report (Xu et al., 1999). Interestingly, the Mucoromycotina MEs in cluster D shared an unusual insertion region (FLxxPG) at the N-terminus preceding the first dinucleotide binding signature motif, which was not found in cluster G or in other yeast and fungal MEs. Moreover, this unusual motif has not been reported for the MEs of other organisms and has no similarity to any known conserved motifs or domains in any proteins available in existing databases. As this peptide was found only in cluster D, it probably represents a region conferring a particular feature to those MEs. However, this should be confirmed experimentally.

**Comparison of 3D structural models of MEs between clusters D and G**

The above analysis of phylogenetic tree and conserved motifs could not provide sufficient information to reveal the divergence between the two closely related MEs of clusters D and G of the Mucoromycotina oleaginous fungi. Thus, 3D structural modelling was utilized to further investigate the location of this particular insertion and its relevance for any possible functional role. The MEs Mucci2_166127 and MucciD_78524 were chosen as representatives of clusters D and G, respectively. The crystal structures of the MEs from many eukaryotic organisms available in the Protein Data Bank (PDB, http://www.pdb.org) (Rose et al., 2011) allowed us to generate structural models of the Mucor enzymes. Based on their sequence information, Mucci2_166127 and MucciD_78524 exhibited 37% and 40% amino acid sequence similarity with the *Ascaris suum* ME (PDB identifier 1O0S) (Rao et al., 2003), respectively. The known structure of the *As. suum* sequence shared one of the highest degrees of sequence identity with the Mucor MEs. According to this level of sequence identity, it was an adequate practical template for the structures of the Mucor MEs and the structure of *Asc. suum* (1O0S) also had a better crystallographic resolution (2.0 Å) than other available known structures of MEs. We therefore exploited the 1O0S structure as a template for modelling the 3D structures of the Mucor MEs. Based on the atomic coordinates of the 1O0S structure, structural models for Mucci2_166127 and MucciD_78524 were constructed by the MODELLER program (Eswar et al., 2006). The generated models of Mucci2_166127 and MucciD_78524, particularly their ribbon diagrams, were compared as illustrated in Fig. 3. Individual folded polypeptides could be separated into four domains (A, B, C and D) as shown in Supplementary Data S4, which is a feature common to most MEs (Xu et al., 1999; Yang et al., 2000; Yang & Tong, 2000; Chang & Tong, 2003). Domain A located at the N-terminus was mostly helical in structure. Domain B contained a central, parallel five-stranded β-sheet surrounded by several helices on both faces. Domain C contained the dinucleotide binding Rossmann fold, with the exception that the third strand was replaced by a short anti-parallel β-structure. Earlier studies have suggested that the cofactor NAD(P)$^+$ is associated with the residue in this region (Xu et al., 1999; Yang & Tong, 2000). The C-terminal domain D consisted of one helix followed by a long extended structure that protruded away from the rest of the monomer similar to observations from previous studies (Xu et al., 1999; Yang & Tong, 2000).

Two consensus regions (dinucleotide binding signature motifs) at aa 207–212 and 353–358 of ME (Mucci2_166127) are conserved across several groups of eukaryotic MEs. It has been documented that these residues are responsible for dinucleotide binding of NAD(P)$^+$ (Yang et al., 2000; Aktas & Cook, 2008). Moreover, the conserved amino acid residues of the divalent metal ion binding site were located at positions similar to those of other MEs and thus might play a role in interactions with divalent metal ions (Yang et al., 2000). Comparison of the two *Mucor* MEs suggests that their catalytic sites are highly similar, even though there was a minor variation in the amino acid sequences of the dinucleotide binding signature motif 2 (Fig. 2b). At the last residue of the motif 2, alanine was found in the motif 2 in Mucci2_166127, whereas glycine was present in MucciD_78524. Furthermore, the structural modelling of the *Mucor* ME of cluster D revealed that the N-terminal insertion region (FLxxPG) at position 159–163 was located in the interface region of domain A (Supplementary Data S4), which is likely involved in dimer formation at the quaternary structure level. Regarding the crystallographic structures of NAD(P)$^+$-dependent MEs, it has been suggested that the dimer interface region provides a continuously positively charged pocket for the binding of negatively charged molecules such as fumarate (activator) and tartronate (competitive inhibitor) (Rao et al., 2003; Karsten et al., 2003). In our model, this insertion region was a protruding site located nearly in the centre of the dimer interface (Fig. 4). In the structure of *Asc. suum* 1O0S, the dimer interface is considered to be an allosteric site allowing strong binding with tartronate, a carboxylic acid analogue to malate and fumarate, via hydrogen bonding with tyrosine [Tyr141(A)].
**Fig. 1.** Dendrogram of yeast and fungal MEs obtained using phylogenetic tree analysis. The numbers over the nodes represent bootstrap coefficients calculated from 1000 replicates. The analysis also included the MEs of *Escherichia coli* and *Synechocystis* sp. to serve as outgroups. The bar indicates evolutionary distance scale.

**Fig. 2.** Amino acid distribution of the conserved residues of MEs. (a) The logo plot was generated according to Crooks *et al.* (2004). The height of each letter is proportional to its frequency at that position and the height of the stack is proportional to information content, measured from its conservation. Each stack is ordered according to frequency, with the most frequent located on top. Hydrophobic residues are shown in black. Polar and neutral amino acids are shown in green and purple, respectively. Acidic and basic amino acids are included in red and blue, respectively. (b) The alignment of amino acid sequences indicates the conserved regions of MEs (see Fig. 1 for species). The numbers in parentheses indicate the number of amino acid residues.
However, the precise function of the insertion region according to the structural modelling of the *Mucor* ME should be further validated experimentally. The key residues contributing to the subunit interaction across the dimer interface included the side-chains of two arginine [Arg105(B) and Arg81(B)] and one glutamine [Gln78(B)] (Rao et al., 2003). A comparison of this putative effector binding pocket between the Mucci2_166127 ME model and the 1O0S template indicated a dramatic change in the configuration of key residues, as shown in Fig. 4. Noticeably, residue Arg105 of the *Ascaris* enzyme structure was substituted with the amide amino acid (Asn121) in the *Mucor* enzyme model. Moreover, we found that the side-chain of phenylalanine (Phe159) (Fig. 4c) was a protruding site, which could possibly interfere in the interactions between effector (such as tartronate or fumarate) and the Asn121 residue. Another side-chain of Tyr157 (in the *Mucor* ME model) was tilted away from the allosteric site compared with that of the homologous residue found in the *Ascaris* template (Tyr141). It is likely that this unusual region in cluster D sequences is involved in dimer organization or in the allosteric properties of these enzymes (Chang et al., 2007). This characteristic might not allow the proper binding of effectors to the pocket, suggesting that it may not require effector binding for its allosteric properties. Previously, it has been reported that the Mucci2_166127 enzyme is NADP$^+$-dependent, and is involved in lipogenesis (Vongsangnak et al., 2012). It is

![Ribbon diagram of the structures of *Mu. circinelloides* MEs. (a) Comparison of structures between Mucci2_166127 (yellow) and MucciD_78524 (light blue), which are members of clusters D and G, respectively. (b) Catalytic binding sites of the MEs are also shown. The dinucleotide binding sites and the divalent metal ion binding sites are highlighted in purple and red, respectively.](image-url)
possible that the discrimination between the characteristics of clusters D and G of the Mucoromycotina MEs, e.g. in terms of enzyme activity and regulation, might result from the observed variation in dimer organization (Supplementary Data S5). With regard to functional structure organization, it has been found that the NADP\(^+\)-dependent isoform of human ME with a stable tetramer behaves in a non-cooperative and non-allosteric manner, whereas the human NAD\(^+\)-dependent isoform performs cooperatively and allosterically (Hsieh et al., 2009). For the latter isoform, the dissociation of subunits at the dimer or tetramer interface reduces the enzyme’s activity by influencing malate binding cooperation and fumarate responsiveness (Hsieh et al., 2009). It is interesting and unclear how this structure of the NADP\(^+\)-dependent ME of Mu. circinelloides facilitates enzyme function. Presumably, each subunit might be able to work independently and not communicate with another, which would make this ME different from other NADP\(^+\)-dependent enzymes. This should be further verified experimentally.

Our study provides additional information of yeast and fungal MEs through an analysis of sequence and structural data. These data not only are useful for the identification of
protein localization and the cofactor preference of MEs, but also lead to a better understanding of the structural characteristics of the key enzyme participating in the generation of NADPH for fatty acid biosynthesis and lipid accumulation in *Mucoromycotina* oleaginous fungi, such as *Mucor* spp. and *Mortierella* spp., which are industrially important micro-organisms.

**ACKNOWLEDGEMENTS**

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission.

**REFERENCES**


Winning, B. M., Bourguignon, J. & Leaver, C. J. (1994). Plant mitochondrial NAD+ -dependent malic enzyme. cDNA cloning,


Edited by: F. Sargent