

DddY is a bacterial dimethylsulfoniopropionate lyase representing a new cupin enzyme superfamily with unknown primary function

Lei Lei, Uria Alcolombri & Dan S Tawfik

Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel.

* Correspondence to: dan.tawfik@weizmann.ac.il

Abstract

Dimethylsulfide (DMS) is released at rates of $>10^7$ tons annually and plays a key role in the oceanic sulfur cycle and ecology. Marine bacteria, algae, and possibly other organisms, release DMS via cleavage of dimethylsulfoniopropionate (DMSP). Different genes encoding proteins with DMSP lyase activity are known belonging to different superfamilies and exhibiting highly variable levels of DMSP lyase activity. DddY shows the highest activity among all reported bacterial lyases yet is poorly characterized. Here, we describe the characterization of recombinant DddY from different marine bacteria. We found that DddY activity demands a transition metal ion cofactor. DddY also shares two sequence motifs with other bacterial lyases assigned as cupin-like enzymes, DddQ, DddL, DddK, and DddW. These cupin motif residues are essential for DddY activity, as for the other cupin DMSP lyases, and all these enzymes are characterized by a common metal-chelator inhibitor (TPEN). Analysis of all sequences carrying these cupin motifs defined a superfamily: Cupin-DLL (DMSP lyases and lyase-like). The DMSP lyase families are sporadically distributed suggesting that DMSP lyases evolved within this superfamily independently along multiple lineages. However, the specific activity levels, genomic context analysis, and systematic profiling of substrate selectivity as described in the accompanying paper, indicate that for only some of these families, most distinctly DddY and DddL, DMSP lyase is the primary, native activity. In other families, foremost DddQ, DMSP lyase seems to be merely a promiscuous activity. The native function of DddQ, and of nearly all members of this newly identified Cupin-DLL superfamily, remains unknown.

Key words: Dimethylsulfide, enzyme superfamily, enzyme promiscuity, substrate profiling.

Abbreviations: DMSP, dimethylsulfoniopropionate; DMS, dimethylsulfide; cupin-DLL, cupin DMSP lyase and lyase-like.

Funding: Financial support by the Estate of Mark Scher, and the Sasson & Marjorie Peress Philanthropic Fund, are gratefully acknowledged. D.S.T. is the Nella and Leon Benozziyo Professor of Biochemistry.

Introduction

Dimethyl sulfide (DMS) is a key regulator of marine life, and possibly also a climate regulator. DMSP (dimethylsulfoniopropionate) is the primary precursor of DMS, and is synthesized mostly by marine phytoplankton (single-cell algae) but also by macroalgae, corals, some angiosperms and bacteria.¹⁻⁵ DMSP probably acts as osmolyte and antioxidant, although its precise physiological role remains unknown, as is the role of DMS release.^{2, 4} Also unknown is the relative contribution of different marine species to the global DMS release, or even of bacterial relative to algal mediated release.⁶ Several putative bacterial DMSP lyase families were identified and assigned as ‘*Ddd*’ (DMSP dependence DMS releasing) genes,⁷ as well as one eukaryote family, dubbed Alma, found in organisms such as algae and corals.⁸

We aimed at better understanding of the enzymology and phylogenetics of most abundant bacterial DMSP lyases. The so-far identified *Ddd* gene families include: DddD, a DMSP CoA-transferase-lyase;^{7, 9} DddP that belongs to the M24 proteinase family;¹⁰ DddQ, DddL, DddW and DddK that have been assigned to the cupin class;¹¹⁻¹⁴ and, DddY whose classification remained unknown so far.^{4, 15-17} The eukaryote Alma DMSP lyase belongs to yet another superfamily (Asp/Glu racemase).⁹ In addition to variable evolutionary origins, these families, and foremost the bacterial enzymes, exhibit highly variable levels of DMSP activity. The reported specific activities seems to range from 0.002 up to 0.028 Units (μM DMS $\text{min}^{-1}\text{mg}^{-1}$ enzyme) for DddQ¹⁸⁻²⁰ up to 675 Units for DddY.^{15, 16} Kinetic parameters are known for only some of these enzymes, but the specific activities relate to k_{cat}/K_M values that may be as low as or even lower than $1 \text{ M}^{-1}\text{s}^{-1}$ for DddQ,¹⁸⁻²⁰ up to $10^6 \text{ M}^{-1}\text{s}^{-1}$ for DddY.^{15, 16} Indeed, DddY, the most active lyase reported to date,^{15, 17, 21} is likely the least characterized one.

To deepen our understanding the enzymology of DMS release, we have further characterized the DddY DMSP lyase family. In doing so, we found that DddY is evolutionary related to DddQ, DddL, DddW and DddK. Together, these 5 families define a new superfamily of enzymes that share the cupin metal site and jellyroll fold. However, the analyses of the kinetic parameters, the substrate specificity (see accompanying paper),

and the genome context, presented here all suggest that while the DddY and DddL represent *bona fide* DMSP lyases, other families, most distinctly DddQ, appear to exhibit DMSP lyase activity as a promiscuous, side-activity. This promiscuous activity stems from shared evolutionary origins and overlapping substrate or/and reaction patterns of members of this newly identified superfamily dubbed the Cupin-DLL (Cupin DMSP lyase and lyase-like) superfamily.

Results

Recombinant DddY's exhibit DMSP lyase activity with k_{cat}/K_M in the range of $10^6 \text{ M}^{-1}\text{s}^{-1}$

To better characterize the DddY family, we collected all putative DddY sequences. We used the previously identified *Alcaligenes* DddY,^{15, 17} to search the NCBI database. After filtering ($\geq 90\%$ coverage and $\geq 30\%$ amino acids identity) and muscle alignment, a phylogenetic tree was built indicating two major clades that are related although with considerable divergence (**Figure 1A**; 32% and 82% average sequence identity compared to AfDddY). The 1st clade included the previously described *Alcaligenes* and *Desulfovibrio* DddY's,^{15, 16} while the 2nd clade included primarily *Ferrimonas* and *Shewanella* genes. Two additional sequences were identified that appear as outgroup of the 2nd clade and notably belong to *Synechococcus*, widely spread marine photosynthetic cyanobacteria (31% and 33% sequence identity compared to AfDddY).

Representatives of these two clades were cloned and over-expressed in *E. coli*, including *Shewanella putrefaciens* CN-32, *Ferrimonas balearica* DSM 9799, *Desulfovibrio acrylicus* (DaDddY), *Alcaligenes faecalis* J481(AfDddY) and *Synechococcus* sp. KORDI-100 DddY genes. *Alcaligenes* DddY was reported to have a periplasmic signal peptide,¹⁷ and a homologous N-terminal region seems to appear in all DddY's. Initially, the wild-type genes including their putative signal peptides were expressed (DNA and amino acid sequences of all constructs are provided as **Supplemental Dataset 1**). Of all tested candidates, DaDddY exhibited the highest activity. However, mass spectrometry of the purified enzyme gave ambiguous results failing to indicate whether the original periplasmic signal peptide was recognized as such in *E. coli*. Thus, the putative mature enzyme sequence

was recloned fused to PelB – a widely used *E. coli* periplasmic signal peptide. This recombinant DaDddY variant was purified yielding the 50kDa mature protein at $\geq 95\%$ purity (**Supplemental Fig. S1A**).

Recombinant DaDddY exhibited maximal activity at pH 8.5, although the pH-rate profile indicated multiple titratable groups (**Supplemental Fig. S1B**) The origins of this double-bell-shaped pH-rate profile are unclear, but a very similar profile was reported for endogenous AfDddY.²² The recombinant AfDddY variant was obtained with its native signal peptide. Both enzymes obeyed the Michaelis-Menten model (**Figure 1B**). DaDddY exhibited a k_{cat}/K_M value of $\sim 1.3 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, and its orthologue, AfDddY, showed ~ 4 -fold lower k_{cat}/K_M of $0.35 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$. These values are highly similar to the values originally reported for the enzymes isolated from their endogenous organisms (**Table 1**). Thus, as indicated by previously reported specific activities, DddY shows the highest catalytic efficiency among all known bacterial DMSP lyases. DddY's catalytic efficiency is also considerably higher than that of Alma algal DMSP lyases ($0.8 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$ for *Emiliana huxleyi* Alma1 and $2.7 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ for *Symbiodinium-A1* Alma1).^{6, 8}

Whilst the 1st clade that includes AfDddY and DaDddY's clearly encompasses highly active DMSP lyases, the identity of the 2nd clade of putative DddY's, including the *Shewanella*, *Ferrimonas* and *Synechococcus* genes, is unclear. Upon expression in *E. coli*, few of the representative genes we tested exhibited weak DMSP lyase activity in crude cell lysates ($\leq 14 \mu\text{M DMS min}^{-1}\text{mg}^{-1}$ total protein in crude lysate; compared to ~ 700 for DaDddY and ~ 400 for AfDddY). However, attempts to purify these enzymes resulted in very low yield and complete loss of activity. Replacing the signal peptide to PelB's did not result in higher yield or activity. Attempts to optimize growth conditions for maximal crude lysate activity gave no improvement and even the initially observed activity failed to reproduce. Lack of activity could relate to misfolding, especially as the yield of soluble protein was low. Inferred ancestors were found to consistently yield proteins with high foldability and stability.²³ We therefore inferred the ancestor of the *Shewanella* clade (**Figure 1A**) and expressed it in *E. coli*. However, the inferred *Shewanella* ancestor expressed poorly and was inactive similar to the *Shewanella putrefaciens* CN-32. Addition

of various metal ions found to be essential for DaDddY (described below) to the growth medium or/and the lysis buffer had no effect either. Hence, at this stage it remains unclear whether the 2nd clade is an integral part of the DddY family although its members are unable to correctly fold in *E. coli*. Alternatively, this clade represents a related enzyme family with some residual, promiscuous DMSP lyase activity. As described below, this ambiguity applies for other members of the Cupin-DLL superfamily that exhibit weak or no DMSP lyase activity upon expression in *E. coli*.

DddY shares cupin motifs with other Ddd families

Having collected a range of DddY sequences, we could identify common sequence motifs, including the cupin motifs that are shared by DddQ, DddL, DddK and DddW^{13, 24, 25}. These cupin motifs could also be identified in DddY. Accordingly, these 5 ddd+ families share two conserved cupin-like sequence motifs (**Figure 2A**). As expected, the two remaining bacterial ddd+ families, dddD and dddP, and the algal Alma DMSP lyases, do not share these motifs as they belong to completely different superfamilies.

The first shared cupin motif comprises 4 entirely conserved residues – HxHxxxxExY (His268, His270, Glu274 and Tyr277 (numbering is for DaDddY). The second motif includes a third entirely conserved histidine, His342. Of the 5 cupin ddd+ families, structures are available for DddQ (PDB 4LA2, Ref.¹⁹; 5JSO and 5JSP, Ref.²⁰; 4B29 and 5CU1), and DddK (PDB 5TG0, Ref.²⁶). The structures indicate that two His residues, one from the 1st motif and another from the 2nd one, and the Glu of the 1st motif, directly chelate the active-site metal ion (His141, Glu145 & His180 in PDB 5CU1). The second His of the 1st motif and its Tyr (His270 and Tyr277 in DaDddY) are also within the active-site and in close proximity to the metal ion (< 4 Å, His125 and Tyr131 in PDB 4LA2; His141 and Tyr147 5CU1; His 150 and Tyr156 in PDB 4B29). Other conserved residues are part of both motifs. For example, a Trp/Tyr at the end of the 2nd motif is also an active-site residue, with the Trp ring's NH placed ~4.5Å from the metal ion (Trp363 in DaDddY; Trp178 in DddQ, PDB 4LA2; Trp195 in PDB 5CU1; Trp204 in PDB 4B29). The location of these residues within the active-site

and close to the metal indicates that they are critical to function and also explains their absolute conservation in DddQ, DddL and DddY families.

Together, these two motifs, and foremost, the above-mentioned 6 active-site residues (**Figure 2A**), define the Cupin-DLL superfamily. Further analysis of this superfamily's gene content beyond the 5 Ddd+ families is provided in the last section of the Results.

DddY demands a transition-metal ion cofactor

Given DddY's sequence assignment as a cupin DMSP lyase, its catalytic activity should be utterly dependent on a transition metal ion.²⁷ Indeed, following incubation with the metal chelator EDTA, DaDddY lost all DMSP lyase activity (**Figure 2B**). However, confusingly, enzymatic activity was restored, and even augmented relative to the untreated enzyme, not only by transition metals (e.g. Co^{2+} , Fe^{2+} , Mn^{2+}) but also by Ca^{2+} that was added as a negative control. It is unlikely that a cupin metal site will bind or be fully active with both an alkali and a transition metal ion. We thus surmised that Ca^{2+} binds at an alternative site that is permissive to both alkali and transition metal ions, while EDTA actually failed to remove the transition metal ion at the cupin site. Indeed, when TPEN, a chelator selective for transition metals, was applied, the DMSP lyase was completely lost and could be fully recovered by addition of transition metal ions, but not by Ca^{2+} (**Figure 2B**). This result confirmed that DddY's activity is dependent on a transition metal cofactor, as observed with all cupin-like enzymes, including DddW and DddQ.^{13, 14, 19} The alternative metal binding site is likely required for maintaining DddY's structure, and may be unique to this family whose size is far larger than any other Cupin ddd+ (**Table 1**).

The conserved motifs are essential for the DMSP cleavage activity

To confirm the role of the most conserved amino acids in the active site of DddY, a series of site-directed mutants to Ala was generated at DaDddY's background. The mutations H268A, H270A, E274A and H342A all showed complete loss of the activity, as reported for other cupin Ddd+ enzymes such as DddW.²⁵ The Y277A mutation retained around 20% of

the wild-type's activity. However, when Tyr277 was replaced by phenylalanine whose side-chain is similar to tyrosine, enzymatic activity was completely lost (**Figure 2C**). The CD spectra of all these mutants were essentially identical to wild-type, suggesting that the mutations perturbed the active-site rather than the enzyme's overall fold (**Supplemental Fig. S2A**). These residues of the two cupin motifs are therefore essential for enzymatic activity, foremost, for the ligation of the active-site metal ion.²⁷ However, Tyr277 that belongs to the first motif presents a puzzle. This tyrosine has been suggested to play a key role in proper positioning of the transition metal ion and the substrate in DddQ's active-site (Tyr131 is PDB 4LA2, 5JSO and 5JSP; Ref. ^{19, 20}).²⁰ However, its substitution to Ala in DaDddY has only a minor effect. Unexpectedly, however, the more conservative substitution to Phe led to complete inactivation, possibly because in the absence of the interacting hydroxyl of Tyr, a hydrophobic moiety in the metal's close vicinity is highly deleterious. To get a clear understanding of the role of this active-site tyrosine, its mutants in DddK, DddW and DddL were examined. Similarly to DddY, the Tyr277 to Ala mutants showed a relatively mild decrease in activity (around 10% residual activity). In DddL and DddW, the Phe mutants showed near-complete loss of activity as in DaDddY, whereas in DddK replacements to either Ala or Phe had similar effect (around 10% residual activity, **see Supplemental Fig. S2B**). Overall, it appears that the Tyr277 of the 1st motif plays a role in DaDddY's DMSP lyase activity, as well is in the other Cupin Ddd+ enzymes, but its role is secondary, certainly in comparison to the metal-ligating residues (His342 in DaDddY, His163 in DddQ, PDB 4LA2).

DddL is potentially a highly active DMSP lyase

The shared cupin motifs, and the role and location of their conserved residues, indicate that DddY, DddQ, DddL, DddK, and DddW belong to the same superfamily. However, as shown for a number of enzyme superfamilies, common evolutionary origin typically results in overlapping activities whereby the native function of one family is observed as promiscuous in related one(s), and vice versa.²⁸⁻³¹ The levels of DMSP lyase activity in Cupin-DLL family members range over 6 orders-of magnitude (**Table 1**). While enzymatic

parameters are widely distributed, the k_{cat}/K_M values within enzyme classes do not vary as widely; foremost, k_{cat}/K_M values of $< 1 \text{ M}^{-1}\text{s}^{-1}$ are highly unlikely to reflect the native substrate (the average k_{cat}/K_M value is $\sim 10^5 \text{ M}^{-1}\text{s}^{-1}$).³² This suggests that whereas DddY's native function is almost certainly as DMSP lyase ($k_{cat}/K_M \approx 10^6 \text{ M}^{-1}\text{s}^{-1}$), DddQ's DMSP lyase activity is likely to be promiscuous ($k_{cat}/K_M < 1 \text{ M}^{-1}\text{s}^{-1}$; **Table 1**). To complete the picture with respect to activity levels, we tested the DMSP lyase activity of members from all five Ddd+ families. Foremost, we sought to determine DddL's activity,³² as to our knowledge, reports of its specific activity, let alone of its kinetic parameters, are not available.

We examined two DddL orthologues: the previously described *Rhodobacter sphaeroides* DddL (RsDddL)¹¹, a DddL gene we identified in the *Thioclava pacifica* genome (TpDddL; 78% amino acid identity to *R. sphaeroides* DddL), and *Sulfitobacter* EE-36 DddL (50% amino acid to RsDddL).¹¹ RsDddL was cloned and expressed in *E. coli* with His-tags at either the N- or C-terminus, or an N-terminal Strep-tag. High DMSP lyase activity was repetitively observed in crude cell lysates, indicating a soluble, correctly folded and active enzyme. However, all attempts to purify RsDddL failed. The C-terminal His-tag construct did not bind to Ni-NTA beads, the N-terminal His-tag construct was insoluble and inactive, while the N-terminal Strep-tag construct bound the resin, but its enzymatic activity was lost upon elution (**Supplemental Fig S3**). Subsequently, TpDddL and *Sulfitobacter* DddL was cloned and expressed with an N-terminal Strep-tag, but showed similar behavior. Various additives were tested, including detergents and metal ions, but enzymatic activity could not be retained during purification. Nonetheless, the specific activity of both DddL's in freshly prepared *E. coli* crude cell lysates, as estimated from its protein concentration and DMSP lyase activity, seems high, in the range of 70 Units (**Supplemental Fig S3**). The kinetics indicated a linear increase of initial rates with DMSP concentrations up to 10 mM suggesting a high K_M and accordingly high k_{cat} . This value of specific activity is probably underestimated, as the soluble fraction is likely to contain misfolded enzyme. Nonetheless, this specific activity is up to 10-fold higher than DddK and DddW's, and is well above 1000-fold higher than DddQ's (**Table 1**).

Genome context suggests that DddQ and DddW's primary activity is not DMSP lyase

In bacterial genomes, enzymes of the same pathway are typically found in gene clusters or even in the same operon. Thus, genome context can provide important hints regarding the enzymatic function of *ddd+* genes. Acrylate, the verified product of all cupin Ddd+ enzymes (**Table 1**), is toxic,^{33, 34} and previous studies indicated proximal genes that relate to acrylate catabolism. These include: *acul*, a zinc/iron dependent alcohol reductase that converts acrylyl-CoA into propanoyl-CoA;^{17, 33} *acuK* (a dehydratase) and *acuN* (a CoA transferase) that can jointly convert acrylate into 3-hydroxypropionate;³⁵ or cytochrome dependence oxi-reductases that could also catabolize acrylate.³⁶

We applied few common tools of genome context analysis, including EFI-GNT,³⁷ STRINGE,³⁸ EASYFIG,³⁹ and RODEO.⁴⁰ Of these, RODEO provided the most systematic results for the examined Ddd+ genes, including the previously studied DMSP CoA-transferase/lyase *dddD* that serves as benchmark for a *bona fide* DMSP lyase.⁹ Accordingly, *dddD* is characterized by proximal *dddB* and *dddc* genes, encoding an iron-containing dehydrogenase and a methyl-malonate semi-aldehyde dehydrogenase-like protein, respectively (**Supplemental Fig S5A**; Refs.^{2, 7}). The product of DddD, 3-hydroxypropionate-CoA, is converted into malonate semi-aldehyde (DddB) and then to acetyl-CoA (DddC).² Accordingly, *dddY*'s genome neighborhoods repetitively contain putative acrylate utilizing genes (Ref ¹⁷; **Supplemental Fig S5B**). Indeed, the bacterium in which DaDddY originally resides (*Desulfovibrio acrylicus*) converts acrylate into propionate;⁴¹ and *Alcaligenes faecalis* M3A strain that contains AfDddY converts acrylate into 3-hydroxypropionate.⁴²

In DddL's case, however, the genome context is not conserved, making it hard to draw a clear conclusion (in 4 out of 15 genomes an *acul*-like zinc containing reductase is present just next to DddL; **Supplemental Fig. S6A**). Similarly, because DddK sequences are only found in several strains of *Pelagibacter ubique* (SAR11),¹² it is difficult to get a systematic prediction based on genome context. In the genome of SAR11, the proximity of enoyl-ACP-reductase, β -ketoacyl-ACP-synthase and β -hydroxydecanoyl-ACP

dehydratase indicate a relation to fatty acid or polyketide biosynthesis.⁴³ (**Supplemental Fig. S6B**)

The genome neighborhoods of both *dddQ* and *dddW* could be consistently derived. For *DddW*, the proximal D-alanyl-D-alanine carboxypeptidase gene (**Supplemental Fig. S6C**) suggests a gene cluster that is involved in bacterial peptidoglycan biosynthesis. The conserved neighborhood of *dddQ*, on the other hand, includes a putative mandelate racemase-like protein and a putative dimethylglycine dehydrogenase (**Figure 3**). The mandelate racemase-like gene exhibits present distinct homology (40% and 35% amino acid identities, respectively) to two recently identified enzymes: *cis*-3-hydroxy-L-proline dehydratase and 4-hydroxyproline betaine 2-epimerase.^{44, 45} These two neighbors of *dddQ*, and the dimethylglycine dehydrogenase-like neighbor, suggest that *DddQ* takes part in the catabolism of proline-betaine or/and hydroxyproline-betaine. Its DMSP lyase activity is therefore likely to be promiscuous, as also indicated by its markedly low activity (**Table 1**) and lack of selectivity (see accompanying paper).

Other putative members of the Cupin-DLL superfamily

To gain further insights into the evolutionary relationships of the cupin DMSP lyases, pfam model PF16867 was used to search against the NCBI protein database and the Tara marine metagenomics database.⁴⁶ Following filtering, the remaining sequences were aligned (361 sequences in total) and an extended profile that encompassed all potential superfamily members was generated (see **Methods** and **Supplemental Dataset 2**). A phylogenetic tree was also built (**Supplemental Fig. S4**). Needless to say that a tree that includes sequences with as low as 15% identity and highly variable lengths (90 - 441 amino acids) is inaccurate. Nonetheless, it provided an initial view of the content of this superfamily and of where the *Ddd+* families map within it, as schematically summarized in **Figure 4**.

The cupin-DLL superfamily is large, and its various clades are widely diverged, suggesting high functional diversity. *DddQ* is by far the largest family. With the exception of *DddW* and *DddK* that are closely related, the other 4 *Ddd+* families are sporadically

distributed throughout the tree, suggesting an independent emergence of DMSP lyases from enzymes with another function, at least for DddL and DddY. However, the vast majority of Cupin-DLLs are obviously not DMSP lyases. However, as exemplified by DddQ, the shared evolutionary origin can give rise to promiscuous DMSP lyase activity.

Cupin-DLL enzymes are inhibited by the metal chelator TPEN and unaffected by Br-DMSP

We have recently reported that 2-bromo-3-(dimethylsulfonio)-propionate (Br-DMSP) is a potent and selective mechanism-based inhibitor of the algal Alma DMSP lyases, and does not inhibit any known bacterial Ddd+ enzymes.⁶ This inhibitor, that covalently modifies the active-site cysteine of the Alma lyases, is a biochemical probe that could detect the presence of Alma DMSP lyase in various marine organisms and environments and assess their contribution to DMS release. We sought to verify that TPEN (**Figure 2B**) is a complementary probe that could identify the cupin DMSP lyases. To this end, DddY, DddL, DddQ, DddK and DddW were incubated with either Br-DMSP or with the transition metal chelator, TPEN. The reaction buffer also contained CaCl₂ to avoid Alma or DddY losing activity (likely due to a structural calcium site; see **Figure 2B**). The Alma DMSP lyases from *Emiliania huxleyi*⁸ and *Symbiodiniome*⁶ were examined as control. All cupin Ddd+ enzymes were all inhibited by TPEN while Br-DMSP had no effect. Conversely, both Alma DMSP lyases were inhibited by Br-DMSP and unaffected by TPEN. Surprisingly, DddP, as a member of the metallo-enzyme superfamily M24 peptidase (see also accompanying paper), showed no inhibition by TPEN; a similar result has been reported in Refs.^{47, 48}, neither EDTA nor 1,10-phenanthroline inhibited DddP's activity. The selective pattern of inhibition by Br-DMSP and TPEN can therefore be applied to identify the presence of the two major classes of DMSP lyases, Alma and cupin-DLL, in different marine environments and species, be they algae or bacteria. This biochemical profiling can complement genomic data to unravel the origins of marine DMS release.

Discussion

The search for DddY's affiliation led to the identification of an enzyme superfamily dubbed Cupin-DLL (**Figure 4**). With the exception of DddD and DddP, all bacterial ddd+ genes implicated as DMSP lyases belong to this newly identified superfamily. We have also identified TPEN as a common inhibitor of the cupin Ddd+ enzymes.

The phylogenetics suggest that DMSP lyases evolved within this superfamily independently along multiple lineages. This phenomenon of parallel evolution is commonly observed in enzyme superfamilies.^{30, 49} Its origins are in shared promiscuous activities - relatively distant superfamily members often share the same promiscuous activity. Such promiscuous activities comprise the starting point for the divergence of new enzymes by turning a latent, promiscuous activity into a physiologically relevant function, initially alongside the enzyme's original function. Indeed, bi-functional, 'generalist' enzymes commonly comprise evolutionary intermediates, although they may persist for long periods as is apparent by the dominance of multifunctional enzymes in extant proteomes.²⁸ As the divergence process proceeds, highly active, selective 'specialist' enzymes evolve, typically by duplication and sub-functionalization of the bi-functional generalist ancestor.⁵⁰

Parallel evolution, as well as convergent evolution as exemplified by the independent evolutionary origins of the algal Alma1 DMSP lyase, also relates to DMSP lysis being a facile reaction,² and foremost to the fact that lysis is initiated by abstraction of a proton from a carbon next to a carboxylate (α -carbon). This is one of the most common steps in enzyme catalysis, and several large enzyme superfamilies share it as the key catalytic step, including the Asp/Glu racemase superfamily to which the Alma DMSP lyases belong. Accordingly, DMSP lyases that belong to the enolase superfamily - where α -proton abstraction is the hallmark, are also highly likely to exist.⁵¹

Further, our view as evolutionary biochemists is that the Ddd+ families likely represent an entire range of evolutionary states - they may be fully diverged DMSP lyase specialists (DddY and DddL), bi-functional intermediates (possibly DddK/W) or enzymes with a different function that only exhibit latent, promiscuous DMSP lyase activity (DddQ in most likelihood). The level of activity is an important parameter with respect to

physiological relevance. However, low specific activity may also be due to suboptimal folding or/and reaction conditions. We thus pursued an independent test of substrate profiling. As described in the accompany paper, it appears that activity levels and substrate selectivity are correlated – the least active enzymes, foremost DddQ, also show the weakest signature of a DMSP tailored active-site.

DddK and DddW reside in the mid-range by both criteria, activity and selectivity, and could therefore be bi-functional enzymes. Enzymes in secondary metabolism tend to exhibit lower catalytic efficiency compared to central metabolism enzymes³² and also exhibit multi-functionality. Indeed, while their DMSP lyase activity of DddK/W is modest ($k_{cat}/K_M \approx 10^3 \text{ M}^{-1}\text{s}^{-1}$), their rates are notably comparable to the rates of DMSP demethylation by the corresponding enzyme (DmdA).²⁶ This comparison is relevant because DMSP is catabolized via two routes, lysis to give DMS as described here, or demethylation to give methylmercapto propionate,^{2, 4, 52} and the demethylation route is actually the dominating one.^{2, 3, 53}

While following the identification of the cupin-DLL superfamily the picture with respect to the DMSP lyases has become somewhat clearer. However, the key remaining question is the primary function of the cupin-DLL superfamily, and specifically of dddQ that also comprises its largest clade (**Figure 4**). Future research may reveal the function of Cupin-DLLs that are not DMSP-lyases, and may thus also shade light on the evolutionary history and function of the ddd+ families.

Material and Methods

Phylogenetic analysis. Hypothetical cupin-like DMSP lyases sequence were collected by using Hidden Markov Model based search using the HMMsearch of the HMMer package.⁵⁴ Briefly, the existing HMM model of DMSP lyase, PF16867, was used to search the NCBI non-redundant protein database and the Tara metagenomics database.⁴⁶ All hits

were collected, aligned, and manually filtered by the presence of the two conserved motifs and minimal length (≥ 90 amino acids). The sequence redundancy in the remaining set was minimized using CD-hit with 70% cutoff. The resulting 361 sequences were aligned by MUSCLE,⁵⁵ followed by minimization of gaps (manual removal of sporadic insertions) to obtain a core alignment of 115 amino acids length (the smallest known Cupin Ddd+ enzymes (DddK's) are 130 amino acids length). A phylogenetic tree were built from this core alignment using the Markov chain Monte Carlo (MCMC) method of the MrBayes program.⁵⁶

Enzyme cloning and mutagenesis. All tested DddY genes (including *DadddY*, *AfdddY*, *Shewanella dddY*, *Ferrimonas dddY*, and *Synechococcus dddY*) were synthesized by Gen9. The *Ruegeria pomeroyi* DSS-3 *dddQ* and *Rhodobacter sphaeroides* 2.4.1 *dddL* genes were kindly provided by Professor Andrew Johnston, University of East Anglia. The synthesized fragments were amplified and the PCR product was digested with *NcoI* and *HindIII*, and then cloned into the expression vector pET28a for expression with C-terminal His tag. These plasmids were also used as the template for site-directed mutagenesis. The *TpdddL* was synthesized by Gen9 and contained an N-terminal Strep-tag and a stop codon before the *HindIII* site. It was cloned into pET28a via the *NcoI* and *HindIII* sites. Point mutations in *dddY*, *TpdddL*, *dddK* and *dddW* were introduced with mutagenesis oligos using SOE-PCR-based approach. The PCR fragments were digested with *NcoI* and *HindIII* and cloned into pET28a. All mutants were verified by DNA sequencing.

Enzyme purification. Enzymes were typically expressed using pET28a plasmid in *E.Coli* BL21 (DE3). Cells were grown for overnight in 5 mL LB medium at 37°C. These cultures (1 mL) were used to inoculate 1 liter LB cultures that were subsequently grown at 37°C to OD_{600nm} of 0.6-0.8. The growth temperature was reduced to 16°C, and enzyme expression was induced with 0.1mM IPTG. Following overnight growth at 16°C, the cells were harvested by centrifugation at 4°C. DddY were purified on Ni-NTA beads. Briefly, cells were re-suspended in 50mL lysis buffer (100mM Tris-HCl pH 8.0, 100mM NaCl, 1mM CaCl₂, 10 mg lysozyme and 10 µg bezonase). Cell suspensions were incubated in ambient temperature for 30 min and sonicated. Lysates were clarified by centrifugation and

loaded on 2 mL Ni-NTA agarose beads (Millipore). Binding was performed at 4°C, beads were washed with 50 mL lysis buffer followed by 100 mL lysis buffer with 35 mM imidazole. The enzyme was eluted with 150 mM imidazole. Fractions were analyzed by SDS-PAGE and activity, combined, and the purified enzyme was concentrated by ultrafiltration (Amicon). For TpDddL with N-terminal Strep-tag, clarified cell lysates were loaded on 2mL Strep-tactin beads (IBA), and the beads were rinsed with 50mL lysis buffer. The protein was eluted with 2.5mM desthiobiotin. Final enzyme concentrations were determined by the BCA assay.

DMSP lyase activity assays. DMS release was measured as previously described.¹⁰ Briefly, freshly prepared 100 mM Tris-HCl pH 8.0 with 100 mM NaCl was used for the enzymatic assays supplemented with 10 mM DMSP as default. The high buffer capacity is critical for this assay, because DMSP as applied as hydrochloride salt, and its cleavage releases protons. Reactions were performed at 30 °C (typically for 5 min,) and terminated by 1000-fold dilution into 30 mL chilled 10mM glycine pH 3.0 in sealed glass vials. Enzyme concentration were typically as follows: *Ehux-Alma1*, 0.3ug/mL; *Sym-Alma1*, 1µg/mL; *Desufovibrio* DddY 20ng/mL; *Alcaligenes* DddY 50ng/mL; DddW, 15µg/mL; DddQ, 100µg/mL; DddK, 8µg/mL. DddL was assayed in crude lysate, at an estimated concentration of 8ug/mL (**Supplemental Fig. S3**). DMS levels were determined using an Eclipse 4660 Purge-and-Trap Sample Concentrator system (OI Analytical) followed by separation and detection using GC-FPD (HP 5890) equipped with RT-XL sulfur column (Restek). All measurements were calibrated using DMS standards.

Metal chelation and complementation. Protein samples were incubated with 1mM EDTA or TPEN, at 30 °C for 1 hr. The chelators were removed by dialysis at 4 °C overnight. Different metal ions were supplemented by adding the corresponding chloride salts to the dialyzed apo-proteins at 2 mM concentration, and incubating at 30 °C for 1 hr. DMSP lyase activity was subsequently measured.

Table 1: Features of the explored DddY and DddL enzymes compared to other cupin-DLL family members

Family	Species	NCBI Accession	Length (amino acids)	Specific activity ($\mu\text{mol min}^{-1} \text{mg enzyme}^{-1}$)	k_{cat}/K_M ($\text{M}^{-1}\text{s}^{-1}$)	K_M (mM)	k_{cat} (s^{-1})	Metal cofactor
DddY	<i>Alcaligenes faecalis</i> J481	E7DDH2.1	401	902±35	$3.5*10^5$	2.56	$0.9*10^3$	Mn ²⁺ (2)
				390 (Ref. ¹⁵)	$2.3*10^5$ (Ref. ¹⁵)	1.41 (Ref. ¹⁵)	$0.3*10^3$ Ref. ¹⁵	
	<i>Desulfovibrio acrylicus</i>	SHJ73420.1	403	1391±76	$1.32*10^6$	0.85	$1.13*10^3$	Mn ²⁺ (2)
				2100 (Ref. ¹⁶)	$4.60*10^6$ (Ref. ¹⁶)	0.45 (Ref. ¹⁶)	$2.07*10^3$ (Ref. ¹⁶)	
<i>Ferrimonas kyonanensis</i>	WP_028114584.1	418	Unclear (1)	N.D.	N.D.	N.D.	N.D.	
<i>Shewanella putrefaciens</i> CN-32	WP_011920089.1	414	Unclear (1)	N.D.	N.D.	N.D.	N.D.	
DddL	<i>Thioclava pacific</i>	WP_051692700.1	232	≥83 (3)	N.D.	N.D.	N.D.	Mn ²⁺
	<i>Rhodobacter sphaeroides</i> 2.4.1	Q3J6L0.1	232	≥70 (3)	N.D.	N.D.	N.D.	Mn ²⁺
DddQ	<i>Ruegeria pomeroyi</i> DSS-3	AAV94883.1	201	$2\sim5*10^{-3}$ (Refs. ^{13, 18})	<20 (Ref. ^{13, 18})	N.D.	N.D.	N.D.
	<i>Ruegeria lacuscaerulensis</i> ITI-1157	D0CY60	192	$2.8*10^{-2}$ (Ref. ²⁰)	0.27 (Ref. ²⁰)	39.1 (Ref. ²⁰)	$1.05*10^{-2}$ (Ref. ²⁰)	Fe ³⁺ (Ref. ²⁰)
DddK	<i>Candidatus Pelagibacter ubique</i> HTCC1062	WP_011281678.1	130	~30 (Ref. ²⁶)	608 (Ref. ²⁶)	5.1 (Ref. ²⁶)	3.1 (Ref. ²⁶)	Ni ²⁺ (Ref. ²⁶)
DddW	<i>Ruegeria pomeroyi</i> DSS-3	WP_011046214.1	152	61.2 Ref. ²⁵	$2.10*10^3$ (Ref. ²⁵)	4.5 (Ref. ²⁵)	17.33 (Ref. ²⁵)	Mn ²⁺ (Ref. ²⁵)

Footnotes:

N.D. not determined

¹ Weak DMSP lyase activity was observed in crude lysates upon over-expression but activity was lost during purification. Attempts to reproduce the lysate activity gave ambiguous results.

² Ca²⁺ is additionally required to obtain maximal activity, probably binding to a structural rather than the cupin site.

³ Specific activity was estimated from the activity and protein concentrations in the crude lysate; see **Supplemental Fig. S3**.

References

1. Stefan M. Sievert, R. P. K., Heide N. Schulz-Vogt (2007) The Sulfur Cycle, *Oceanography* 20.
2. Curson, A. R., Todd, J. D., Sullivan, M. J., and Johnston, A. W. (2011) Catabolism of dimethylsulphonioacetate: microorganisms, enzymes and genes, *Nature reviews. Microbiology* 9, 849-859.
3. Reisch, C. R., Moran, M. A., and Whitman, W. B. (2011) Bacterial Catabolism of Dimethylsulphonioacetate (DMSA), *Frontiers in microbiology* 2, 172.
4. Johnston, A. W., Green, R. T., and Todd, J. D. (2016) Enzymatic breakage of dimethylsulphonioacetate—a signature molecule for life at sea, *Current opinion in chemical biology* 31, 58-65.
5. Curson, A. R., Liu, J., Bermejo Martinez, A., Green, R. T., Chan, Y., Carrion, O., Williams, B. T., Zhang, S. H., Yang, G. P., Bulman Page, P. C., Zhang, X. H., and Todd, J. D. (2017) Dimethylsulphonioacetate biosynthesis in marine bacteria and identification of the key gene in this process, *Nature microbiology* 2, 17009.
6. Alcolombri, U., Lei, L., Meltzer, D., Vardi, A., and Tawfik, D. S. (2016) Assigning the algal source of dimethylsulfide using a selective lyase inhibitor, *ACS Chemical Biology*.
7. Todd, J. D., Rogers, R., Li, Y. G., Wexler, M., Bond, P. L., Sun, L., Curson, A. R., Malin, G., Steinke, M., and Johnston, A. W. (2007) Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria, *Science* 315, 666-669.
8. Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D. S., and Vardi, A. (2015) MARINE SULFUR CYCLE. Identification of the algal dimethyl sulfide-releasing enzyme: A missing link in the marine sulfur cycle, *Science* 348, 1466-1469.
9. Alcolombri, U., Laurino, P., Lara-Astiaso, P., Vardi, A., and Tawfik, D. S. (2014) DddD is a CoA-transferase/lyase producing dimethyl sulfide in the marine environment, *Biochemistry* 53, 5473-5475.
10. Todd, J. D., Curson, A. R., Dupont, C. L., Nicholson, P., and Johnston, A. W. (2009) The dddP gene, encoding a novel enzyme that converts dimethylsulphonioacetate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi, *Environmental microbiology* 11, 1376-1385.
11. Curson, A. R., Rogers, R., Todd, J. D., Brearley, C. A., and Johnston, A. W. (2008) Molecular genetic analysis of a dimethylsulphonioacetate lyase that liberates the climate-changing gas dimethylsulfide in several marine alpha-proteobacteria and *Rhodobacter sphaeroides*, *Environmental microbiology* 10, 757-767.
12. Sun, J., Todd, J. D., Thrash, J. C., Qian, Y., Qian, M. C., Temperton, B., Guo, J., Fowler, E. K., Aldrich, J. T., Nicora, C. D., Lipton, M. S., Smith, R. D., De Leenheer, P., Payne, S. H., Johnston, A. W., Davie-Martin, C. L., Halsey, K. H., and Giovannoni, S. J. (2016) The abundant marine bacterium *Pelagibacter* simultaneously catabolizes dimethylsulphonioacetate to the gases dimethyl sulfide and methanethiol, *Nature microbiology* 1, 16065.
13. Todd, J. D., Curson, A. R., Kirkwood, M., Sullivan, M. J., Green, R. T., and Johnston, A. W. (2011) DddQ, a novel, cupin-containing, dimethylsulphonioacetate lyase in marine roseobacters and in uncultured marine bacteria, *Environmental microbiology* 13, 427-438.

14. Todd, J. D., Kirkwood, M., Newton-Payne, S., and Johnston, A. W. (2012) DddW, a third DMSP lyase in a model Roseobacter marine bacterium, *Ruegeria pomeroyi* DSS-3, *The ISME journal* 6, 223-226.
15. de Souza, M. P., and Yoch, D. C. (1995) Purification and characterization of dimethylsulfoniopropionate lyase from an alcaligenes-like dimethyl sulfide-producing marine isolate, *Applied and environmental microbiology* 61, 21-26.
16. van der Maarel, M. J. E. C., Aukema, W., and Hansen, T. A. (1996) Purification and characterization of a dimethylsulfoniopropionate cleaving enzyme from *Desulfovibrio acrylicus*, *FEMS Microbiology Letters* 143, 241-245.
17. Curson, A. R., Sullivan, M. J., Todd, J. D., and Johnston, A. W. (2011) DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of Proteobacteria, *The ISME journal* 5, 1191-1200.
18. Alcolombri, U., Elias, M., Vardi, A., and Tawfik, D. S. (2014) Ambiguous evidence for assigning DddQ as a dimethylsulfoniopropionate lyase and oceanic dimethylsulfide producer, *Proceedings of the National Academy of Sciences of the United States of America* 111, E2078-2079.
19. Li, C. Y., Wei, T. D., Zhang, S. H., Chen, X. L., Gao, X., Wang, P., Xie, B. B., Su, H. N., Qin, Q. L., Zhang, X. Y., Yu, J., Zhang, H. H., Zhou, B. C., Yang, G. P., and Zhang, Y. Z. (2014) Molecular insight into bacterial cleavage of oceanic dimethylsulfoniopropionate into dimethyl sulfide, *Proceedings of the National Academy of Sciences of the United States of America* 111, 1026-1031.
20. Brummett, A. E., and Dey, M. (2016) New Mechanistic Insight from Substrate- and Product-Bound Structures of the Metal-Dependent Dimethylsulfoniopropionate Lyase DddQ, *Biochemistry* 55, 6162-6174.
21. Yoch, D. C., Ansele, J. H., and Rabinowitz, K. S. (1997) Evidence for Intracellular and Extracellular Dimethylsulfoniopropionate (DMSP) Lyases and DMSP Uptake Sites in Two Species of Marine Bacteria, *Applied and environmental microbiology* 63, 4625.
22. de Souza, M. P., and Yoch, D. C. (1995) Comparative Physiology of Dimethyl Sulfide Production by Dimethylsulfoniopropionate Lyase in *Pseudomonas doudoroffii* and *Alcaligenes* sp. Strain M3A, *Applied and environmental microbiology* 61, 3986-3991.
23. Trudeau, D. L., Kaltenbach, M., and Tawfik, D. S. (2016) On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins, *Molecular biology and evolution* 33, 2633-2641.
24. Sullivan, M. J., Curson, A. R., Shearer, N., Todd, J. D., Green, R. T., and Johnston, A. W. (2011) Unusual regulation of a leaderless operon involved in the catabolism of dimethylsulfoniopropionate in *Rhodobacter sphaeroides*, *PLoS one* 6, e15972.
25. Brummett, A. E., Schnicker, N. J., Crider, A., Todd, J. D., and Dey, M. (2015) Biochemical, Kinetic, and Spectroscopic Characterization of *Ruegeria pomeroyi* DddW--A Mononuclear Iron-Dependent DMSP Lyase, *PLoS one* 10, e0127288.
26. Schnicker, N. J., De Silva, S. M., Todd, J. D., and Dey, M. (2017) Structural and Biochemical Insights into Dimethylsulfoniopropionate Cleavage by Cofactor-Bound DddK from the Proliferative Marine Bacterium *Pelagibacter*, *Biochemistry* 56, 2873-2885.
27. Dunwell, J. M., Purvis, A., and Khuri, S. (2004) Cupins: the most functionally diverse protein superfamily?, *Phytochemistry* 65, 7-17.
28. Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective, *Annual review of biochemistry* 79, 471-505.

29. Roodveldt, C., and Tawfik, D. S. (2005) Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily, *Biochemistry* **44**, 12728-12736.
30. Glasner, M. E., Gerlt, J. A., and Babbitt, P. C. (2006) Evolution of enzyme superfamilies, *Current opinion in chemical biology* **10**, 492-497.
31. Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase, *Biochemistry* **38**, 4252-4258.
32. Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters, *Biochemistry* **50**, 4402-4410.
33. Todd, J. D., Curson, A. R., Sullivan, M. J., Kirkwood, M., and Johnston, A. W. (2012) The *Ruegeria pomeroyi* acul gene has a role in DMSP catabolism and resembles yhdH of *E. coli* and other bacteria in conferring resistance to acrylate, *PLoS one* **7**, e35947.
34. Curson, A. R., Burns, O. J., Voget, S., Daniel, R., Todd, J. D., McInnis, K., Wexler, M., and Johnston, A. W. (2014) Screening of metagenomic and genomic libraries reveals three classes of bacterial enzymes that overcome the toxicity of acrylate, *PLoS one* **9**, e97660.
35. Todd, J. D., Curson, A. R., Nikolaidou-Katsaraidou, N., Brearley, C. A., Watmough, N. J., Chan, Y., Page, P. C., Sun, L., and Johnston, A. W. (2010) Molecular dissection of bacterial acrylate catabolism--unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production, *Environmental microbiology* **12**, 327-343.
36. Mikoulińska, O., Akimenko, V., Galouchko, A., Thauer, R. K., and Hedderich, R. (1999) Cytochrome c-dependent methacrylate reductase from *Geobacter sulfurreducens* AM-1, *European journal of biochemistry* **263**, 346-352.
37. Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., and Whalen, K. L. (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks, *Biochimica et biophysica acta* **1854**, 1019-1037.
38. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic acids research* **43**, D447-452.
39. Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011) Easyfig: a genome comparison visualizer, *Bioinformatics* **27**, 1009-1010.
40. Tietz, J. I., Schwalen, C. J., Patel, P. S., Maxson, T., Blair, P. M., Tai, H. C., Zakai, U. I., and Mitchell, D. A. (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape, *Nature chemical biology* **13**, 470-478.
41. van der Maarel, M. J. E. C., van Bergeijk, S., van Werkhoven, A. F., Laverman, A. M., Meijer, W. G., Stam, W. T., and Hansen, T. A. (1996) Cleavage of dimethylsulfoniopropionate and reduction of acrylate by *Desulfovibrio acrylicus* sp. nov, *Archives of Microbiology* **166**, 109-115.
42. Ansele, J. H., Pellechia, P. J., and Yoch, D. C. (1999) Metabolism of acrylate to beta-hydroxypropionate and its role in dimethylsulfoniopropionate lyase induction by a salt marsh sediment bacterium, *Alcaligenes faecalis* M3A, *Applied and environmental microbiology* **65**, 5075-5081.
43. Massengo-Tiasse, R. P., and Cronan, J. E. (2009) Diversity in enoyl-acyl carrier protein reductases, *Cellular and molecular life sciences : CMLS* **66**, 1507-1517.

44. Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., Bonanno, J. B., Hillerich, B. S., Seidel, R. D., Babbitt, P. C., Almo, S. C., Sweedler, J. V., Gerlt, J. A., Cronan, J. E., and Jacobson, M. P. (2013) Discovery of new enzymes and metabolic pathways by using structure and genome context, *Nature* 502, 698-702.
45. Zhang, X., Kumar, R., Vetting, M. W., Zhao, S., Jacobson, M. P., Almo, S. C., and Gerlt, J. A. (2015) A unique cis-3-hydroxy-l-proline dehydratase in the enolase superfamily, *Journal of the American Chemical Society* 137, 1388-1391.
46. Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmiento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans, c., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., and Bork, P. (2015) Ocean plankton. Structure and function of the global ocean microbiome, *Science* 348, 1261359.
47. Kirkwood, M., Le Brun, N. E., Todd, J. D., and Johnston, A. W. (2010) The dddP gene of *Roseovarius nubinhibens* encodes a novel lyase that cleaves dimethylsulfonylpropionate into acrylate plus dimethyl sulfide, *Microbiology* 156, 1900-1906.
48. Wang, P., Chen, X. L., Li, C. Y., Gao, X., Zhu, D. Y., Xie, B. B., Qin, Q. L., Zhang, X. Y., Su, H. N., Zhou, B. C., Xun, L. Y., and Zhang, Y. Z. (2015) Structural and molecular basis for the novel catalytic mechanism and evolution of DddP, an abundant peptidase-like bacterial Dimethylsulfonylpropionate lyase: a new enzyme from an old fold, *Molecular microbiology* 98, 289-301.
49. Elias, M., and Tawfik, D. S. (2012) Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases, *The Journal of biological chemistry* 287, 11-20.
50. Soskine, M., and Tawfik, D. S. (2010) Mutational effects and the evolution of new protein functions, *Nature reviews. Genetics* 11, 572-582.
51. Gerlt, J. A., Babbitt, P. C., Jacobson, M. P., and Almo, S. C. (2012) Divergent evolution in enolase superfamily: strategies for assigning functions, *The Journal of biological chemistry* 287, 29-34.
52. Reisch, C. R., Moran, M. A., and Whitman, W. B. (2008) Dimethylsulfonylpropionate-dependent demethylase (DmdA) from *Pelagibacter ubique* and *Silicibacter pomeroyi*, *Journal of bacteriology* 190, 8018-8024.
53. Reisch, C. R., Stoudemayer, M. J., Varaljay, V. A., Amster, I. J., Moran, M. A., and Whitman, W. B. (2011) Novel pathway for assimilation of dimethylsulphonylpropionate widespread in marine bacteria, *Nature* 473, 208-211.
54. Eddy, S. R. (2011) Accelerated Profile HMM Searches, *PLoS computational biology* 7, e1002195.
55. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic acids research* 32, 1792-1797.
56. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, *Systematic biology* 61, 539-542.

Figure Legends

Figure 1. DddY phylogeny and kinetic characterization.

(A) A maximum-likelihood tree of identified putative DddY sequences indicates two clades highlighted in pink and blue. . Three genes were characterized from the top clade (indicated in arrows), and the ancestral node for the *Shewanella* sub-clade (blue dot) was inferred, yet none of these 4 proteins showed consistent DMSP lyase activity. The bottom clade, however, encompasses highly active DMSP lyases. The Complete gene names and entry numbers are given in **Supplemental dataset 1**. (B) Determination of the kinetic parameters of two recombinant DddY's from the blue clade: *Desulfovibrio* and *Alcaligenes* DddY's (DaDddY, AfDddY, respectively). Rates were determined in a buffer containing 1 mM CaCl₂ under initial rates conditions (see Methods). Data points represent the average specific activity for 3 independent measurements and the error bars represent the S.D. values. Shown is a direct fit to the Michaelis-Menten model ($R^2 \geq 0.97$). The resulting kinetic parameters are presented in **Table 1**.

Figure 2. DddY is a cupin-like enzyme.

(A) Sequence alignment of the two most conserved regions of the reported cupin Ddd+ families and DddY. The resulting sequence Logo is presented on top. Marked in asterisks are the 6 most conserved active-site residues. (B) Incubation of DaDddY with either EDTA or TPEN (2 mM) resulted in complete loss of activity. Addition of Ca²⁺ restored activity after EDTA treatment, whereas addition of transition metal ions restored activity after treatment with both chelators. (C) The activity of various mutants of DaDddY. Shown is relative activity compared to wild-type ($\mu\text{M DMS/min}$, at 100ng/mL enzyme; mutants were tested at 100ng/mL). Data points in panels B and C represent the average activity in 3 independent measurements and the error bars represent the S.D. values.

Figure 3. Our search identified 8 dddQ-like genes in sequenced genomes. Genome context analysis was performed by RODEO⁴⁰, annotation of each genes were obtain from the original genome annotation, and manually checked. The genes encoding orthologues are highlighted with the same colour. The result suggests that *dddQ* gene is associated with the catabolism of proline-betaine or/and hydroxyproline-betaine. Shown are the results.

Figure 4. A schematic phylogenetic tree of the Cupin-DLL family (based on the phylogenetic tree presented in **Supplemental Figure S4**). Overall, 8 clades could be identified that include the 5 known Ddd+ families, whereby DddK and DddW belong to the same clade. Of the 4 Ddd+ clades, only DddY and DddL seem to be specialized DMSP lyases. The primary function of the other Ddd+ clades, and foremost of DddQ, and of the 4 additional clades (marked arbitrarily as 1-4), remains unknown.

Figure 5. Cupin Ddd+ DMSP lyases are selectively inhibited by the metal chelator TPEN. All tested enzyme were incubated with 10uM Br-DMSP or 1mM TPEN (100mM Tris, 100mM NaCl, 1 mM CaCl₂, pH 8.0). The residual DMS release rates were tested after 1 hour incubation at ambient temperature. Shown is the relative DMSP lyase activity compared to the untreated enzyme. The Alma DMSP lyases from *Emiliania huxleyi* and *Symbiodinium* are marked as E-hux and sym, respectively.

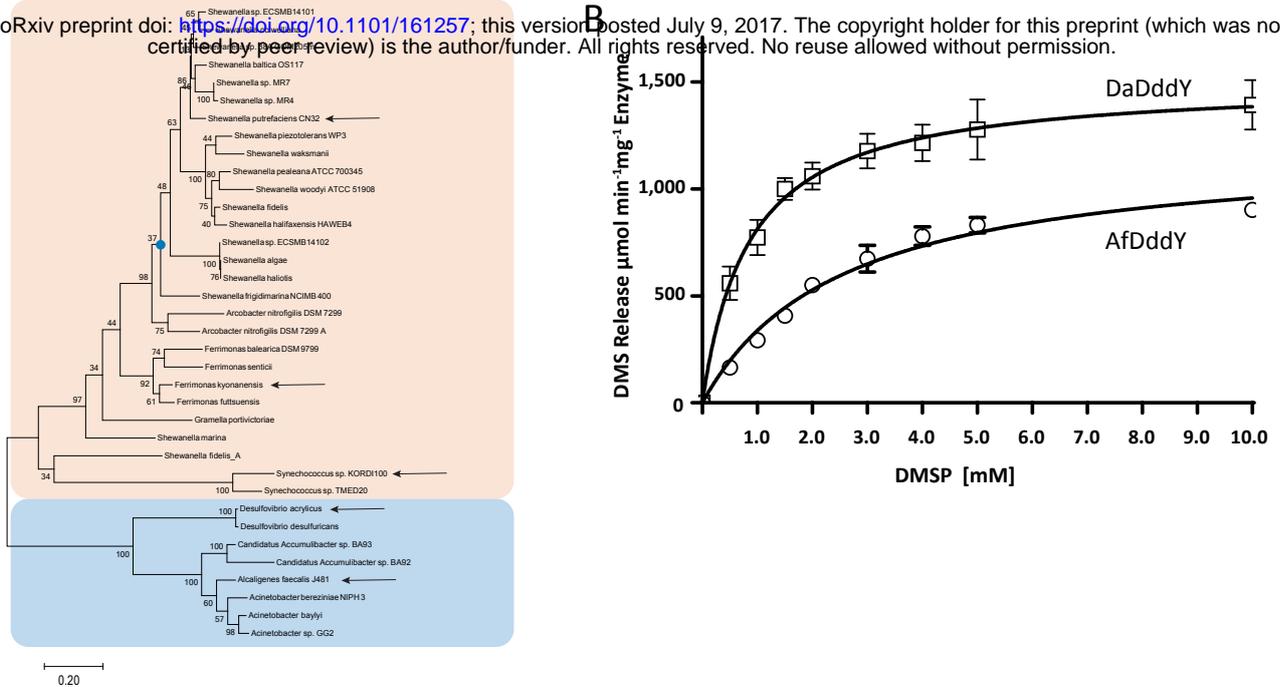


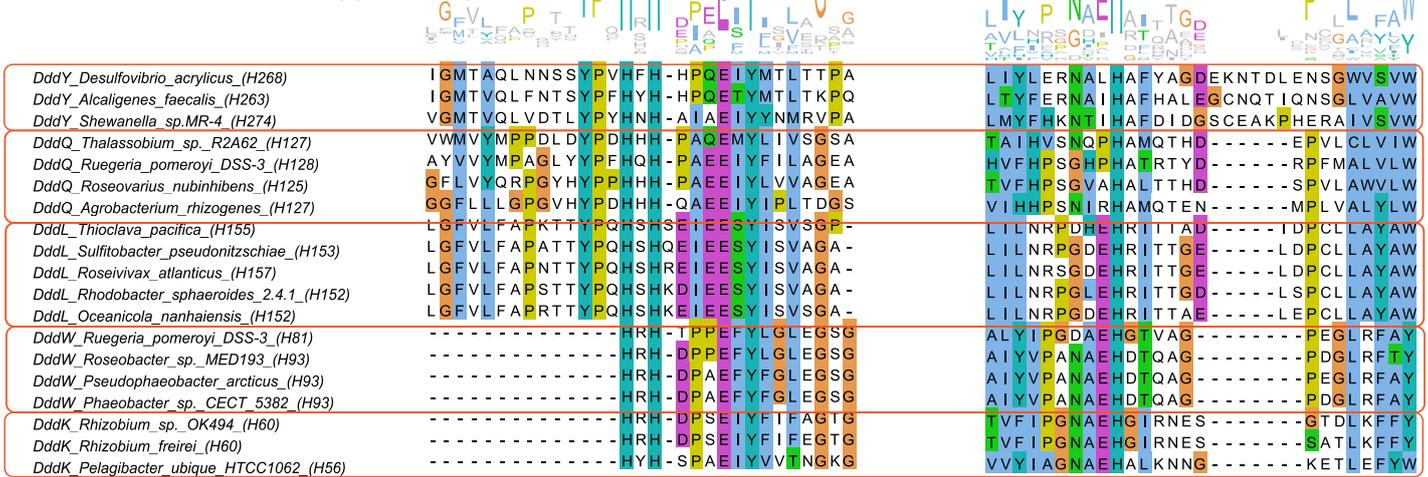
Figure 1. DddY phylogeny and kinetic characterization.

(A) A maximum-likelihood tree of identified putative DddY sequences indicates two clades highlighted in pink and blue. Three genes were characterized from the top clade (indicated in arrows), and the ancestral node for the *Shewanella* sub-clade (blue dot) was inferred, yet none of these 4 proteins showed consistent DMSP lyase activity. The bottom clade, however, encompasses highly active DMSP lyases. The Complete gene names and entry numbers are given in Supplemental dataset1. (B) Determination of the kinetic parameters of two recombinant DddY's from the blue clade: *Desulfovibrio* and *Alcaligenes* DddY's (DaDddY, AfDddY, respectively). Rates were determined in a buffer containing 1 mM CaCl_2 under initial rates conditions (see Methods). Data points represent the average specific activity for 3 independent measurements and the error bars represent the S.D. values. Shown is a direct fit to the Michaelis-Menten model ($R^2 \geq 0.97$). The resulting kinetic parameters are presented in **Table 1**.

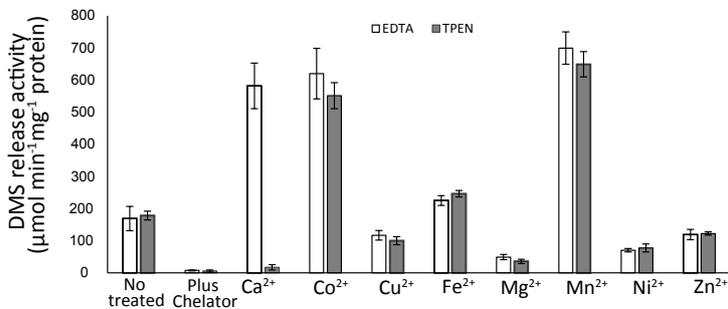
Motif 1

Motif 2

A bioRxiv preprint doi: <https://doi.org/10.1101/161257>; this version posted July 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



B



C

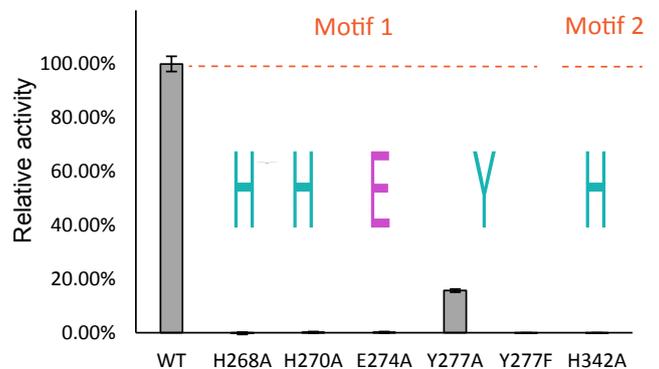


Figure 2. DddY is a cupin-like enzyme.

(A) Sequence alignment of the two most conserved regions of the reported cupin Ddd+ families and DddY. The resulting sequence Logo is presented on top. Marked in asterisks are the 6 most conserved active-site residues. (B) Incubation of DaDddY with either EDTA or TPEN (2 mM) resulted in complete loss of activity. Addition of Ca²⁺ restored activity after EDTA treatment, whereas addition of transition metal ions restored activity after treatment with both chelators. (C) The activity of various mutants of DaDddY. Shown is relative activity compared to wild-type (μM DMS/min, at 100ng/mL enzyme; mutants were tested at 100ng/mL). Data points in panels B and C represent the average activity in 3 independent measurements and the error bars represent the S.D. values.

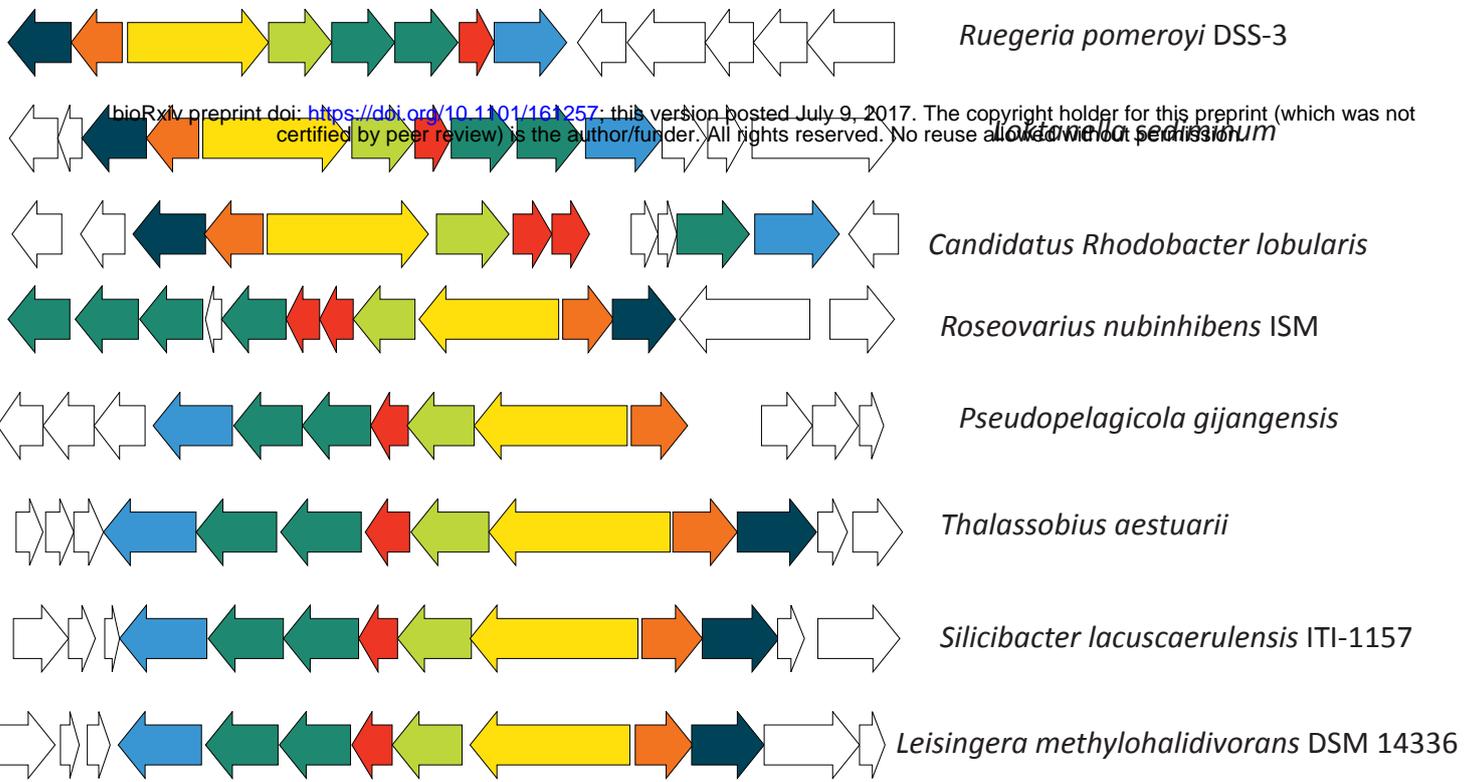


Figure 3. Our search identified 8 dddQ-like genes in sequenced genomes. Genome context analysis was performed by RODEO40, annotation of each genes were obtain from the original genome annotation, and manually checked. The genes encoding orthologues are highlighted with the same colour. The result suggests that *dddQ* gene is associated with the catabolism of proline-betaine or/and hydroxyproline-betaine. Shown are the results.

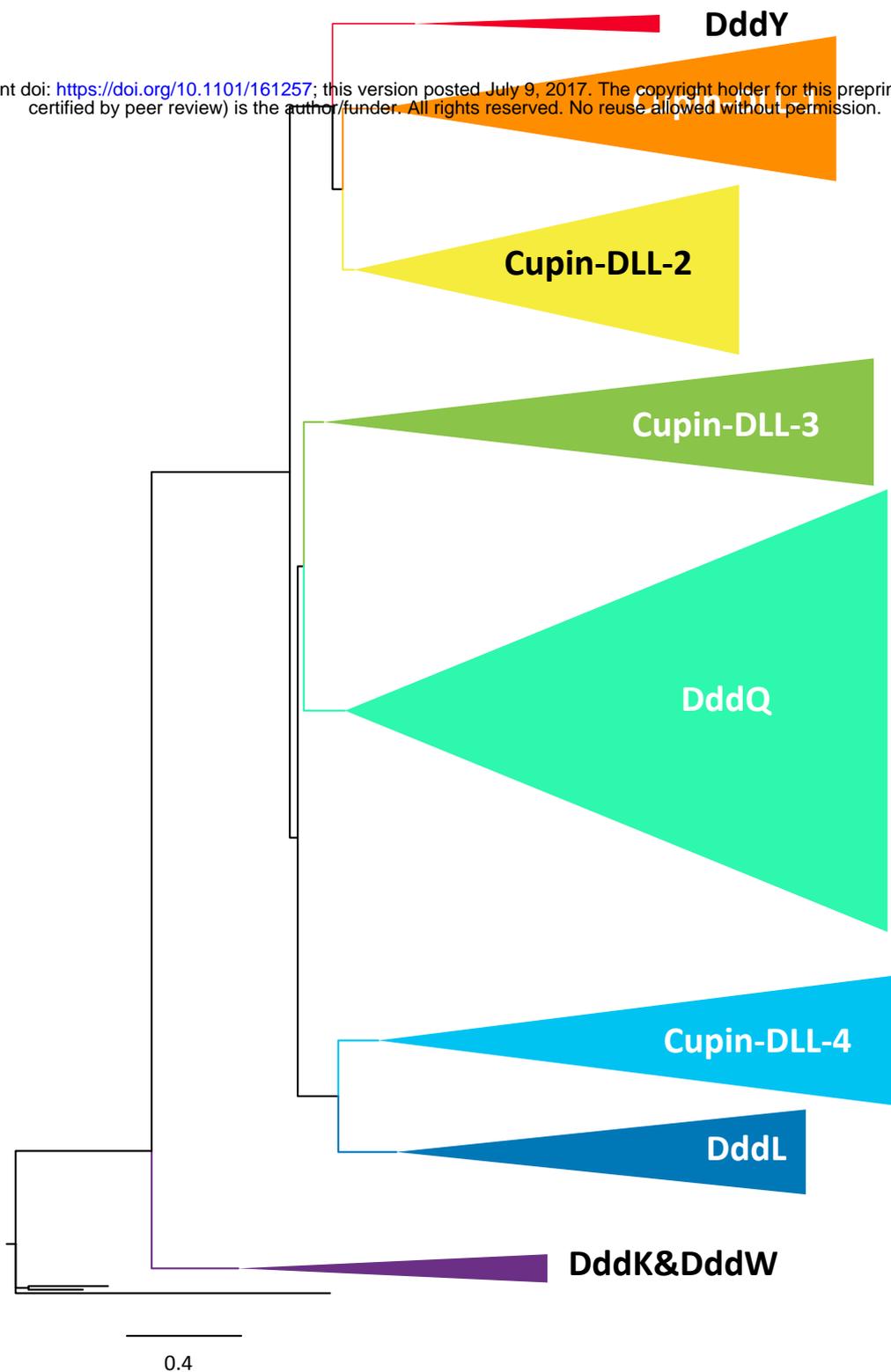


Figure 4. A schematic phylogenetic tree of the Cupin-DLL family (based on the phylogenetic tree presented in **Supplemental Figure S4**). Overall, 8 clades could be identified that include the 5 known Ddd+ families, whereby DddK and DddW belong to the same clade. Of the 4 Ddd+ clades, only DddY and DddL seem to be specialized DMSP lyases. The primary function of the other Ddd+ clades, and foremost of DddQ, and of the 4 additional clades (marked arbitrarily as 1-4), remains unknown.

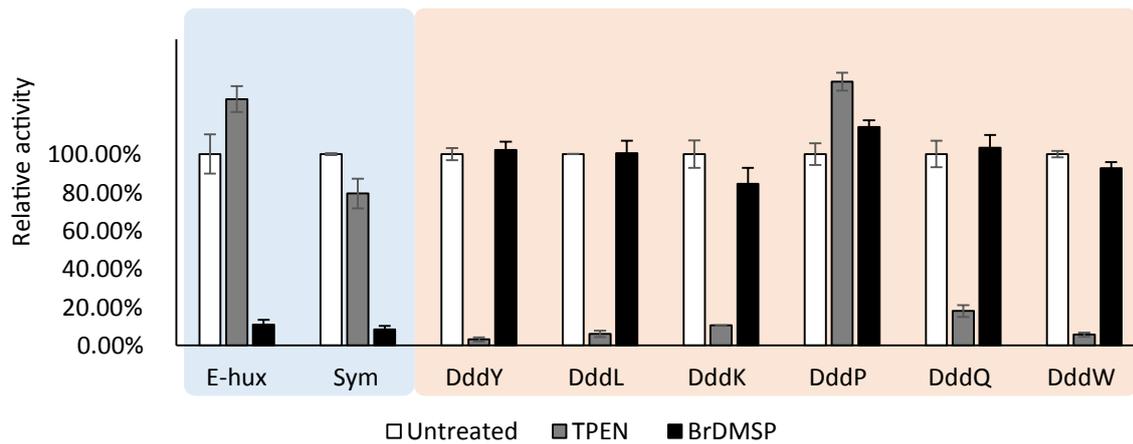


Figure 5. Cupin Ddd+ DMSP lyases are selectively inhibited by the metal chelator TPEN. All tested enzymes were incubated with 10 μ M Br-DMSP or 1mM TPEN (100mM Tris, 100mM NaCl, 1 mM CaCl₂, pH 8.0). The residual DMS release rates were tested after 1 hour incubation at ambient temperature. Shown is the relative DMSP lyase activity compared to the untreated enzyme. The Alma DMSP lyases from *Emiliana huxleyi* and *Symbiodinium* are marked as E-hux and sym, respectively.