Multiplexed Assessment of Promiscuous Non-Canonical Amino Acid Synthase Activity in a Pyridoxal Phosphate-Dependent Protein Family

Anna Zmich, Lydia J. Perkins, Craig Bingman, Justin F. Acheson, and Andrew R. Buller*

ABSTRACT: Pyridoxal phosphate (PLP)-dependent enzymes afford access to a variety of non-canonical amino acids (ncAAs), which are premier building blocks for the construction of complex bioactive molecules. The vinylglycine ketimine (VGK) subfamily of PLP-dependent enzymes plays a critical role in sulfur metabolism and is home to a growing set of secondary metabolic enzymes that synthesize γ-substituted ncAAs. Identification of VGK enzymes for biocatalysis faces a distinct challenge because the subfamily contains both desirable synthases and lyases that break down ncAAs. Some enzymes have both activities, which may contribute to pervasive mis-annotation. To navigate this complex functional landscape, we used a substrate multiplexed screening approach to rapidly measure the substrate promiscuity of 40 homologs in the VGK subfamily. We found that enzymes involved in transsulfuration are less likely to have promiscuous activities and often possess undesirable lyase activity. Enzymes from direct sulfuration and secondary metabolism generally had a high degree of substrate promiscuity. From this cohort, we identified an exemplary γ-synthase from Caldicellulosiruptor hydrothermalis (CahyGS). This enzyme is thermostable and has high expression (∼400 mg protein per L culture), enabling preparative-scale synthesis of thioether containing ncAAs. When assayed with L-allylglycine, CahyGS catalyzes a stereoselective γ-addition reaction to afford access to a unique set of γ-methyl-branched ncAAs. We determined high-resolution crystal structures of this enzyme that define an open-close transition associated with ligand binding and set the stage for future engineering within this enzyme subfamily.

KEYWORDS: biocatalysis, sequence similarity network analysis, methionine analogs, noncanonical amino acids, competition reactions

The number of catalogued putative enzymes from Nature has exploded during the modern genomic era.¹ These sequences are a rich source for the identification of enzymes that have useful properties for biocatalysis. For example, pyridoxal phosphate (PLP)-dependent enzymes are preeminent for the manipulation of amino acids, and many have found industrial use.²−⁴ The broad utility of non-canonical amino acids (ncAAs) in medicinal chemistry⁵,⁶ and chemical and synthetic biology⁷−⁹ has spurred a range of biocatalytic strategies to synthesize them.¹⁰−¹⁴ Of particular utility are enzymes that can activate an amino acid donor molecule to react with a range of substrate partners in a convergent and stereoselective manner. Some native PLP-dependent enzymes satisfy these criteria and operate on preparative scales to yield a variety of substituted ncAAs.¹⁵−²⁵ When native enzymatic activity is too low, directed evolution can be used to improve the activity of convergent, complexity-building ncAA synthases.²⁶−³⁵ Recently, PLP-dependent enzymes from outside central metabolism were found to catalyze γ-substitution en route to a variety of structurally diverse ncAAs.³⁶−³⁸ Enzymes that catalyze γ-substitution (called γ-synthases) operate through a common vinylglycine ketimine (VGK) intermediate (Figure 1A). Currently, it is not known which γ-synthases have desirable substrate promiscuity or which ones suffer from parasitic reaction pathways like hydrolysis. Beyond mechanism, synthetic utility benefits from a blend of properties. Ideally, an enzyme would have high soluble expression in Escherichia coli, could drive reactions to high yields, and could be thermostable and crystallizable to enable downstream engineering.

Because enzymes from both primary and secondary metabolism have the potential to operate on preparative scales, it is helpful to consider a wide set of enzymes that catalyze γ-substitution.³⁶ These enzymes play key roles in the principle metabolic pathways for sulfur metabolism and are found throughout all three domains of life. The key γ-S bond

Received: June 1, 2023
Revised: August 7, 2023
of l-methionine (Met) can be formed through a variety of pathways, the simplest of which is direct sulfuration (DS).

The DS pathway is predominant among bacteria and features a single VGK enzyme with several names such as homocysteine (Hcy) synthase, 37 O-acetylhomoserine (OAHS) sulfhydrylase, 38,39 and OAHS aminocarboxypropyltransferase. 40 Here, we refer these enzymes as Hcy γ-synthases. These enzymes use an activated homoserine analog to form the VGK intermediate, which is electrophilic at Cγ (Figure 1B). 40 Nucleophilic attack by hydrogen sulfide or methylmercaptan followed by a series of tautomerizations then furnishes Hcy or 33-35 γ-acetylhomoserine (OAHS), which is then subject to a forward transsulfuration (TS) pathway (red) uses a cystathionine γ-synthase (CGS) to synthesize cystathionine (Cth), while reverse TS (gray) uses a cystathionine γ-lyase (CGL) to break down Cth. (D) Secondary metabolic pathways are shown with a solid green arrow and feature many VGK enzymes.

Figure 1. Metabolic pathways involving enzymes from the vinylglycine ketimine (VGK) subfamily. (A) γ-Synthases produce γ-substituted amino acids, and γ-lyases break down γ-substituted amino acids. (B) The direct sulfuration (DS) pathway is shown with a blue solid arrow and uses a homocysteine γ-synthase. (C) The forward transsulfuration (TS) pathway (red) uses a cystathionine γ-synthase (CGS) to synthesize cystathionine (Cth), while reverse TS (gray) uses a cystathionine γ-lyase (CGL) to break down Cth. (D) Secondary metabolic pathways are shown with a solid green arrow and feature many VGK enzymes. CGS enzymes (Figure 1). 36 This similarity causes ambiguity about the native enzyme function, which is compounded by functional promiscuity. For example, E. coli CGS has promiscuous CGL activity and Fub7 has retained Hcy γ-synthase activity. 34,42 Other metabolic functions of γ-lyases have been suggested through the observation of Hcy and L-canavanine (Cnv) γ-lyase activity. 43,44 It is not known whether these γ-lyases play a role in secondary metabolism, if γ-lyase activity is common among VGK-utilizing enzymes 33-35,46 or whether any of these myriad functions can be segregated through bioinformatics.

Here, we disclose our efforts to deconvolute activities across the VGK subfamily and to identify exemplary enzymes for biocatalysis. SSN analysis of ~38,000 homologs led to an initial grouping from which we experimentally analyzed 40 homologs. While there have been advances in bioinformatic tools to analyze enzyme sequences and structures, 31,47-49 experimental assignment of enzyme function across protein families still largely relies on detailed measurements of single-substrate reactions with purified enzymes. 50-52 Often, observation of activity with a single substrate is not enough to infer broader synthetic utility. We used a combination of single-substrate and substrate-multiplexed screens to acquire diagnostic information about the γ-lyase and γ-synthase properties of each enzyme in our panel. From these screens, we identified a sequence motif that correlates to TS activity. Further, these assays revealed several well-behaved and promiscuous enzymes. We identified a CGL with activity on a range of γ-functionalized amino acids. We also identified a γ-synthase from Caldicellulosiruptor hydrothermales (CahyGS) that is likely involved in an unknown secondary metabolic pathway and performs well as an in vitro biocatalyst. We used this enzyme for the preparative-scale synthesis of methionine analogs. Excitingly, this enzyme can
react with L-allylglycine and catalyze a stereoselective $\gamma$-addition to yield a new class of $\gamma$-branched ncAAs. X-ray crystal structures of CahyGS reveal conformational changes associated with ligand binding and set the stage for future enzymological and synthetic efforts.

## RESULTS

### Sequence Similarity Network Analysis of the VGK Subfamily.

Previous studies showed that enzymes within the VGK subfamily are catalytically promiscuous and perform both $\gamma$-synthase and $\gamma$-lyase reactions.\(^{42,44,53,54}\) To capture a broad cohort of sequences in this subfamily and enable diagnostic distinctions between these enzymes, we performed a series of BLAST queries against previously characterized homologs from both primary and secondary metabolism, along with four predicted CGS homologs from thermophilic organisms, a CGL with promiscuous Hcy $\gamma$-lyase activity, and an MGL (see the Supporting Information for further detail). The resulting $\sim$38,000 unique sequences were organized via a sequence similarity network (SSN) from the Enzyme Function Initiative–Enzyme Similarity Tool (EFI-EST).\(^{41}\) To ease analysis and reduce the computational load, sequences sharing the same length and at least 70% sequence identity were compressed into 5263 nodes prior to any clustering analysis. This large number of distantly related homologs is indicative of the broad functional diversity within the VGK subfamily. Within the VGK subfamily, we observed that bona fide MGLs are resolved and phylogenetically distinct at an alignment score of 125 ($\sim$50% sequence ID) (Figure 2A). However, previously characterized enzymes in the primary metabolic DS pathway and secondary metabolism were not separable from one another through SSN analysis, even at high alignment cutoffs (Figure 2B,C).

We chose 40 diverse homologs across the SSN for in vitro characterization. This library size allowed for each homolog to be screened in duplicate along with associated controls in a standard 96-well plate. We chose to not include any MGLs in the library, as these enzymes appeared well resolved in the SSN (Figure 2A). The downstream goal of this research is to identify well-behaved (i.e., high expressing, thermostable, substrate promiscuous) enzymes for biocatalysis. Thus, we over-sampled sequences from clusters with known enzymes in secondary metabolism and from thermophilic bacteria (23 total). To confirm that MGLs were indeed phylogenetically distinct, we screened the homolog library for the breakdown of Met into $\alpha$KB (2), ammonia, and methanethiol. We did not observe any evidence of MGL activity and confirmed that MGLs have distinct activities within the VGK subfamily (Figure S1).

### Identification of Homologs with Cystathionine $\gamma$-Lyase Activity.

Detailed study of the *E. coli* CGS (EcCGS) and *Salmonella* CGS showed that these enzymes have promiscuous CGL activity in vitro, although this is not thought to serve a physiological role.\(^{42,55,56}\) Curiously, CGS and CGL enzymes share identical active site architectures, which has led to the hypothesis that the reaction specificity is controlled by substrate supply in an in vivo setting.\(^{36}\) To assess the level of CGL activity across a wider set of enzymes, we assayed each enzyme in our homolog library for the breakdown of Cth (1, Figure 3). Because assays with purified enzyme are not

![Figure 2.](image-url) Sequence similarity network (SSN) analysis of the vinylglycine ketimine (VGK) subfamily. Forty homologs were chosen to be screened, and they are denoted by diamonds. (A) 37,793 sequences were compressed into 5263 nodes that share the same length and >70% identity. An alignment score of 125 ($\sim$50% ID) shows that the Met $\gamma$-lyases separate into a distinct cluster, but other functions appear interspersed based on annotations. (B) 13,073 sequences (1479 nodes) of the Hcy $\gamma$-synthase and secondary metabolic $\gamma$-synthase cluster from panel (A) at an alignment score of 176 ($\sim$70% ID). (C) 3879 sequences (500 nodes) of the Hcy $\gamma$-synthase and secondary metabolic $\gamma$-synthase cluster from panel (B) at an alignment score of 187 (>70% ID). Despite this high degree of stringency, unique catalytic functions do not resolve into separate clusters.

![Figure 3.](image-url) Assessment of cystathionine (Cth) $\gamma$-lyase activity among select homologs. Reactions were run using 1.125 mM Cth (red) and 25 $\mu$M enzyme at pH 8.0 for 1 h in triplicate. Substrate and product quantitation was performed by UPLC-MS. $\alpha$KB is shown in gray.
practical at this throughput, we conducted whole-cell assays using overexpressed enzyme in E. coli BL21(DE3). Production of αKB (2), as well as depletion of 1, was measured after 1 h by UPLC-MS. No activity was observed in negative controls containing a whole cell that over-express PLP-dependent enzymes from outside the VGK subfamily (an aminolevulinic acid synthase from Rhodospirillum centenum and tryptophan synthase β-subunit (2B9) from Pyrococcus furiosus).27 Notably, the E. coli host cells used for this study still contain the native CGS gene, but no background γ-lyase activity was observed, presumably due to feedback inhibition of CGS expression in rich media (Figure S2).27−29 However, overexpression of EcCGS results in 2 formation, indicative of promiscuous γ-lyase activity (Figure S2). Across the 40 homologs, we observed 17 enzymes with CGL activity, 12 of which showed high (>70%) conversion (Figure S2). Three enzymes had high conversion of 1 but only low levels of 2, suggestive of another pathway for 1 breakdown. Subsequent analysis showed evidence of Hcy formation, indicative of Cth β-lyase (CBL) activity. Indeed, one of these enzymes was originally annotated as a CBL (Table S1 and Figure S2). The remaining 23 enzymes from the panel showed no reaction with 1.

Interpretation of such “negative” data from a single experiment is tenuous, as many factors might prevent observation of activity. For example, enzymes may be true CGLs but do not express well in E. coli or may not fold properly. Or these enzymes may simply have some other functions. We validated the results from the CGL screen with a trio of purified enzymes corresponding to the three principle phenotypes in triplicate (high, low, and no conversion, Figure 3) with standard curves from single ion retention measurements (Figures S3 and S4). The promiscuous CGL activity
with EcCGS was readily observed with purified enzyme. The most active γ-lyase from the screen, Thermobifida fusca (ThfuCGL), showed high activity in vitro (100% conversion in 1 h, 45 turnovers). The homolog from C. hydrothermalis (Cahy) showed no reaction with I. This enzyme expressed well in E. coli and had the characteristic yellow color of PLP incorporation, so the lack of CGL activity that was observed when assayed in vitro is indicative that the enzyme has some other functions (Figure 3).

Screening for γ-Synthase Activity Reveals Homologs with Different Substrate Specificities. We next screened for activity in the synthetic direction. Most γ-synthases from the VGK subfamily catalyze γ-substitution with OAHS (3) as the pro-electrophile, including secondary metabolic homologs. Notably, E. coli and other γ-proteobacteria use O-succinyl-L-homoserine to reach the VGK intermediate but also have low levels of promiscuous activity with 3.46 We assayed the 40 homologs for γ-synthase activity with Cys (1a) in competition with other thiol nucleophiles that have diverse physical properties (Figure 4A,B). This substrate multiplexed screening (SUMS) approach offers several distinct advantages compared to single-substrate measurements. In SUMS unimolecular transformations, the ratio of the products is directly proportional to the ratio of the catalytic efficiencies. For bimolecular reactions, however, complex kinetic phenomena can break the pseudo-first-order relationship between catalytic efficiency and specificity. When substrates are placed in direct competition, the ratios of the products are a direct reflection of an enzyme’s ability to discriminate between two or more substrates, irrespective of the underlying mechanism. In contrast, comparison of relative rates of product formation from single-substrate reactions provides only indirect information about specificity. These multiplexing approaches have seen sporadic use to characterize enzyme function but have not seen widespread adoption. We hypothesized that the specificity of bona fide CGS enzymes would be apparent by the preferential accumulation of I when assayed in competition. Such specificity has been observed previously in single-substrate reactions with a Thermotoga maritima Hcy γ-synthase from the DS pathway. This enzyme reacts with bisulfide to yield Hcy but has no activity with Cys, reflecting the specificity required for in vivo function.

We avoided the use of bisulfide and methanethiol, which are the native DS substrates, because these compounds are volatile in protic environments and fermentative routes already exist to access their products (Hcy and Met). Further, Hcy γ-synthases are promiscuous with a variety of thiol nucleophiles (but not Cys). Nucleophile promiscuity was also observed with the secondary metabolic enzyme Fub7, as it reacts with bisulfide. We measured promiscuous activities with a range of well-behaved and chemically diverse thiols that might give access to diverse thioether containing ncAAs. Whole-cell reactions were run with 3 and the mixture of five thiol nucleophiles (including Cys). Reactions were allowed to proceed for 1 h before measuring product formation by UPLC-MS. We observed no activity in negative controls where the native EcCGS gene was present. Overexpression of EcCGS enabled observation of a low level of activity with the nonnative pro-electrophile 3.

In accordance with our multiplexing hypothesis, the SUMS data showed that EcCGS has a clear kinetic preference for reactivity with the native nucleophile 1a to form 1 (Figure 4C). To accumulate this product, the expressed EcCGS enzyme must have comparatively slow CGL activity when run under these conditions, consistent with the biological function of the enzyme. Across the wider panel of homologs, we observed activity with 32 of the 40 homologs (Figure 4C). Two additional homologs had activity like that of EcCGS: only 1 production with no γ-lyase activity. Ten homologs had both γ-synthase and γ-lyase activity in this screen. Among these homologs with both modes of activity, six homologs produced predominantly 1.
The remaining homologs showed low reactivity with 2-phenylethylmercaptan (2-PEM, 1d). Last, two enzymes decomposed 3 without forming a new amino acid; each enzyme had high CGL activity in our previous screen. Seventeen homologs had no γ-lyase activity, including the three bona fide secondary metabolic γ-synthases, and instead exhibited γ-synthase activity with 1d. We observed a surprisingly consistent specificity profile among these 17 enzymes. Because this pattern was also observed with homologs from known secondary metabolic pathways, this specificity pattern alone cannot be used to determine whether these γ-synthases belong to a primary DS metabolic pathway or in an as-yet uncharacterized secondary metabolic pathway. We again validated SUMS results using three purified enzymes that correspond to the main phenotypes we observed for the γ-lyase activity screen with standard curves from single ion retention measurements (Figure 4D and Figure S3). All three product profiles matched the results from the whole-cell screen. In vitro results confirmed that the native nucleophile 1a was preferred for EcCGS, albeit at low levels with a non-native pro-electrophile and high catalyst loading. CahyGS showed clear specificity for 1d. ThfuCGL had high promiscuous γ-lyase activity with 3, as shown by the accumulation of 2.

**Multiplexed Assessment of Bimolecular Substrate Promiscuity.** As a further test for catalytic diversity within the homolog library, we assayed for activity with different pro-electrophiles (Figure 5A). Elimination of γ-homoserine (Hse, 4) and γ-canavanine (Cnv, 5) has been previously reported for enzymes of this subfamily.39,45,66 We also wanted to test the ability of enzymes to eliminate the dimethyl sulfoxide analog of Met, S-Me-Met (6). 1 undergoes an O-N acyl shift at basic pH and deacetylates, whereas 6 is water-stable and easy to synthesize. Because the same reactive VGK intermediate is generated regardless of which pro-electrophile, we reasoned that the product specificity profile of each enzyme would be invariant to the amino acid used to enter the catalytic cycle. Therefore, we ran substrate multiplexed assays with the set of four pro-electrophiles and five thiol nucleophiles simultaneously (double SUMS, Figure 5A).

Results from this whole-cell double-multiplexed assay overall were in good agreement with previous experiments (Figure S5). The product profiles for γ-substitution products were similar to the single pro-electrophile assay. However, 10 homologs with γ-lyase and γ-synthase activity preferred 1a as a nucleophile to produce 1, except for one homolog that preferentially reacted with 1d. This result differs slightly from the single SUMS and may reflect complex kinetic phenomena as enzymes react with the substrate milieu. The pro-electrophile apparent specificity was measured by monitoring depletion of each pro-electrophile substrate. The majority, 24 of 34 active enzymes, showed specificity for 3. However, there was a remarkable degree of pro-electrophile promiscuity among a small set of catalysts (Figure S5). Several γ-lyases showed promiscuous activity with each of the alternative pro-electrophiles, and four had high αKB production (>1 mM) in a reaction mixture with a large set of nucleophiles (Figure S5). Additionally, we observed γ-synthase activity with Cnv (5) in the double SUMS assay for two homologs that showed no activity in the single SUMS.

We validated the double-multiplexed screening results with the same trio of purified enzymes in the previous two screens with standard curves from single ion retention measurements (Figure S4). With the addition of three more pro-electrophiles, we observed that EcCGS had a different double SUMS profile from the single SUMS. The enzyme reacted well with the non-native pro-electrophile 5, and noticeably, more 2 was formed (Figure 5B). Curiously, the substitution product with N-acetyl cysteamine (SNAC, 2b) was the predominant γ-synthase product, not the native 1 product that dominated in the initial screens (Figure 5C). To further interrogate this unusual pattern of reactivity, we performed a time course experiment that showed that 1 accumulates at early timepoints, but product reentry enables its elimination to 2, while other substitution products persist (Figure S6). Although the data in Figure 4D and SC were collected with the same enzyme concentration, we hypothesize that the higher pro-electrophile concentration enabled the EcCGS to completely react with Cys under the double-SUMS conditions and begin using the product Cth as a pro-electrophile in a promiscuous CGL reaction. Consequently, 2b begins to accumulate, thereby accounting for the apparent differences in the specificity in the single and double SUMS validations. The product profiles for ThfuCGL and CahyGS were similar between the single and double SUMS with both whole cell and purified enzyme. These data confirm that ThfuCGL is a prolific degrader of γ-substituted amino acids, as it had γ-lyase activity on all four pro-electrophiles. CahyGS showed reactivity with all four pro-electrophiles, but the preferred substrate was 3, which is the native pro-electrophile for most γ-synthases within the VGK subfamily.

**Rapid Categorization of Enzymes in the VGK Subfamily.** Combined, we acquired biochemical data on 34 of the 40 homologs we tested from the VGK subfamily (Table S1). Because of catalytic promiscuity, it is not possible to definitively distinguish between CGSs, CGLs, and secondary metabolic homologs from these data alone. However, we combined this information with additional bioinformatic analysis to suggest the function of most enzymes studied here (see the Supporting Information and Figures S7–S10). From these studies, we observed that enzymes from TS tend to have less activity on substrate analogs (i.e., high specificity for Cys) and are more likely to have adventitious lyase activity. Therefore, it may be helpful for future biocatalysis endeavors to explicitly identify and avoid these enzymes. While SSN is powerful, it is cumbersome to repeat, and we therefore searched for a sequence signature that correlates to our experimental observations of TS activity. We identified a simple active site motif “E325-S326-L327” (E. coli numbering) that appears to be characteristic of the TS enzymes. Alternatively, enzymes either with direct sulfuration activity or that are from secondary metabolism often contain an “A367-N368-V369” (CahyGS numbering) sequence in this region, although more diverse sequences may be found as well (Figure S9). These motifs are easy to identify through simple alignments, without SSN analysis, and may serve as guideposts for subsequent searches for new activity within the VGK subfamily.

**Synthetic Utility of CahyGS.** CahyGS has many desirable characteristics for biocatalysis. The enzyme has high expression in a standard shake flask (~400 mg protein L⁻¹ culture) and can be purified by heat-treatment at 65 °C for 15 min.

To test for synthetic utility, we performed preparative-scale reactions using heat-treated lysate or whole-cell preparations of the catalyst for each of the nucleophiles used in the synthase screens. 3 was used as the pro-electrophile, and reactions were carried out on a 1 mmol scale for the production of γ-.
substituted ncAAs (mechanism found in Figure S11). Each new ncAA was isolated via reverse phase flash chromatography in varying yields (Figure 6). We attempted synthesis of 1 using

\[
\begin{align*}
\text{O} & \quad \text{N} \\
\text{H} & \quad \text{O} \\
100 \text{ mM} & \quad \text{25 mM} \\
O-\text{acetyl-L-homoserine} & \quad S-\text{nucleophile} \\
\quad \text{CahyGS} & \quad \text{whole cell}\frac{a}{b} \\
\quad \text{heat treated lysate}\frac{b}{c} & \quad \text{OH} \\
\end{align*}
\]

Figure 6. Biocatalytic preparative-scale reactions of Met analogs with CahyGS. One millimole γ-substitution preparative-scale reactions with CahyGS for the production of Met analogs. Catalyst loadings were either 1% w/v wet whole cell weight (\(\frac{a}{b}\)) or 1% w/v of a heat-treated and clarified lysate (\(\frac{b}{c}\)). Reactions were run overnight at 37 °C. Isolated yields are provided for the products from all nucleophiles tested in SUMS, except for Cys.

1a. However, these reactions went to low conversion, consistent with the biological discrimination against the 1a nucleophile, and we did not isolate the small amount of 1 formed. Previous synthetic routes to these products often begin from the disulfide and use reducing metal conditions to generate the thiol prior to in situ alkylation under cryogenic conditions. The biocatalytic route using CahyGS avoids the need to reduce a starting material and can be performed under mild conditions. The preparative-scale reactions with whole cells tended to give higher yields when compared to reactions using heat-treated lysates. Interestingly, 2b had a higher yield than 2d, even though the 1d nucleophile was the preferred substrate in SUMS, indicating some kinetic parameters of reactivity in competition are not universal predictors of activity with single substrates when assayed for high yields.

To further probe the biocatalytic potential of this enzyme, we screened for γ-addition products with l-allylglycine (8) as the pro-electrophile. Activity with 8 has been reported for EcCGS, yielding γ-methylcystathionine, and follows subtly different mechanism when compared to γ-substitution (Figure 7A). Instead of an enzyme-catalyzed elimination, there is a tautomerization that uses an unknown acid, potentially a solvent, to protonate the terminal methyl group to generate a methylvinylglycine ketimine. EcCGS-catalyzed attack of Cys into this more hindered electrophile yielded the γ-methyl analog of Cth, but the absolute configuration of the product was unknown.54

We found that CahyGS was also able to bind the l-allylglycine electrophile (Figure S12) and catalyzed a productive γ-addition reaction with four of the five S-nucleophiles (1a was again rejected). Excited by this finding, we performed preparative whole-cell reactions and isolated products with yields comparable to those of the corresponding γ-substitution reactions. This reaction generates a new stereocenter at the γ-position, and the alkyl thiols reacted with high diastereoselectivity, with all but the product with 2-mercaptopypyridine (3e) giving only a single diastereomer by 1H-NMR. We performed Marfey’s derivatization for each ncAA isolated and observed excellent stereochemical purity but were unable to resolve 3e. The absolute stereochemical configuration of the cyclopentanethiol product (3c) was determined by X-ray crystallography, revealing a 2S,4R configuration (Figure 7B). This stereochemistry corresponds to attack by the thiolate nucleophile into the re-face of the methylvinylglycine ketimine intermediate. While some of these preparative-scale yields are modest, subsequent reactivity could be improved through reaction optimization or directed evolution.

Structural Characterization of a Model γ-Synthase. To facilitate future engineering with this promising biocatalyst, we pursued the X-ray crystal structure of CahyGS. This enzyme crystalized readily and yielded two different structures corresponding to open and closed conformations (Figure 8). The open structure was solved to 2.3 Å resolution and had a clear internal aldimine in the active site (Figures S13 and S14 and Table S2). The closed structure was solved to 1.5 Å resolution in the presence of MES buffer (Figures S13 and S14 and Table S2). The MES formed a set of H-bonds to Arg403 which likely correspond to the carboxylate binding motif for external aldimine formation. This binding is associated with a large-scale conformational rearrangement, including the closure of a substrate channel by the 9 Å motion of a helix into a closed state. These data reinforce the role of protein dynamics in catalysis in the VGK subfamily and provide clear views of the enzymes to guide future engineering efforts.

DISCUSSION

Multiplexed Screening Reveals Substrate Promiscuity Trends in the VGK Subfamily. The assignment of an enzyme’s native function is a prominent challenge in enzyme discovery. Often, arduous kinetic characterization is conducted with purified enzyme to ascertain \(k_{cat}/K_m\) values for multiple substrates and therefore infer an enzyme’s specificity. In contrast, competition experiments are operationally simple to conduct and yield more direct information on specificity. Here, we show that competition information from SUMS can be used to rapidly screen the promiscuity of a series of PLP-dependent enzymes in the VGK-subfamily that all operate through a common mechanism. Through a series of three screens, we acquired detailed information on 34 homologous proteins, including both γ-lyases and γ-synthases. Functional groupings of these enzymes were non-obvious from sequence information and SSN analysis alone. By combining biochemical data with genome neighborhood information, we were able to propose putative functions for most enzymes (Table S1 and Figure S8).
Previously, the identity of the active site residues and the presence of catalytic promiscuity within the CGS and CGL enzymes led to the hypothesis that the enzymes are synonymous in some organisms. Instead, the direction of the reaction is determined by substrate concentrations. While this scenario may indeed occur, our data show unequivocally that there are enzymes, such as ThfuCGL, that cannot catalyze a productive CGS reaction. We also observe a case of a CGS (from Thermothelomyces thermophilus) that, under the conditions here, shows low but measurable CGS activity but no detectable γ-lyase activity with Cth. Hence, future studies will be needed to reveal how distal sequence elements influence CGS vs CGL activity and therefore direction within the transsulfuration pathway.

Biocatalytic Access to ncAAs. One of the benefits of rapid scanning of enzyme promiscuity is the identification of starting points for biocatalysis. Fold type II PLP-dependent enzymes, such as tryptophan synthase, have been engineered to access a myriad of β-substituted ncAAs. These efforts have benefited from thermostable parent enzymes and detailed mechanistic information that enabled identification of alternative substrates that react to yield β-branched products. When initial activities were low, structural information was used to guide directed evolution. A comparable effort for γ-synthase enzymes could open access to γ-substituted amino acids, with a suitable synthase to be found.

The data reported here suggest that primary metabolic CGS enzymes are poor candidates for biocatalytic applications. These enzymes generally have high substrate specificity for Cys and promiscuous CGL activity that can decompose products after they are made. These enzymes that were often misannotated would therefore be difficult to avoid. We identified a simple “ESL” sequence motif that flanks the active site and is associated with bona fide function as a CGS or CGL

Figure 7. γ-Addition into L-allylglycine catalyzed by CahyGS. (A) Proposed γ-addition mechanism. (B) One millimole γ-addition preparative-scale reactions with CahyGS for the production of γ-methyl Met analogs. Catalyst loading was 1% wet whole cell weight. Products from all nucleophiles tested in SUMS, except for Cys, were isolated with varying yields. The d.r. was determined via NMR, and e.r. was determined via Marfey’s derivatization. The small-molecule X-ray crystal structure of the γ-addition product with the cyclopentanethiolate nucleophile shows a 2S,4R configuration.
When ambiguity is present, high-throughput assays are needed. The assignment of catalytic activities of newly discovered enzymes is challenging from sequence information alone. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.3c02498.

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.3c02498. Detailed experimental procedures, protein and DNA sequences, including $^1$H NMR spectra for all compounds, LC/MS spectra, and supporting figures and methods (PDF)

Crystallographic data for 3c (CIF)

**CONCLUSIONS**

The assignment of catalytic activities of newly discovered enzymes is challenging from sequence information alone. When ambiguity is present, high-throughput assays are needed to rapidly deconvolute enzyme function. Here, we have shown how information from direct competition assays can answer challenging questions in enzymology and lead to the discovery of well-behaved enzymes for basic science and biocatalysis. In particular, the thermostable and promiscuous CahyGS is an exemplary catalyst for the production of both branched and linear $\gamma$-substituted ncAAs.

**ASSOCIATED CONTENT**

◆ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.3c02498.

Figure 8. Structural analysis of CahyGS. Open (gray) and closed (blue) conformation of CahyGS. The internal aldimine (E(Ain)) structures were solved for both conformations at 2.3 and 1.5 Å resolutions. The closed conformation has a helix shifted by 9.4 Å, encapsulating the bound 2-ethanesulfonic acid (MES) buffer molecule to Arg403 in the active site.

**AUTHOR INFORMATION**

**Corresponding Author**

Andrew R. Buller – Department of Biochemistry and Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin S3706, United States; orcid.org/0000-0002-9635-4844; Email: arbuller@wisc.edu

**Authors**

Anna Znich – Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin S3706, United States

Lydia J. Perkins – Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin S3706, United States

Craig Bingman – Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin S3706, United States; orcid.org/0000-0002-3073-5089

Justin F. Acheson – Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin S3706, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.3c02498

**Funding**

This work was supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation and the NIH DP2-GM137417 to A.R.B and an NIH Biotechnology Training Grant T32-GM008349 to A.Z. The purchase of the Bruker D8 VENTURE Photon III X-ray diffractometer was partially funded by NSF Award #CHE-1919350 to the UW–Madison Department of Chemistry. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. The use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We thank Brian Fox’s lab for the generous use of their FPLC and members of Buller lab for their input throughout the process and their detailed editing of the manuscript. The small-molecule structure of [3c][HCl$\cdot$H$_2$O was performed by Dr. Ilia A. Guzei, Molecular Structure Laboratory of the Department of Chemistry.

**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLP</td>
<td>pyridoxal phosphate</td>
</tr>
<tr>
<td>CSL</td>
<td>cystathionine synthase-like</td>
</tr>
<tr>
<td>VGK</td>
<td>vinylglycine ketimine</td>
</tr>
<tr>
<td>SSN</td>
<td>sequence similarity network</td>
</tr>
<tr>
<td>DS</td>
<td>direct sulfuration</td>
</tr>
<tr>
<td>TS</td>
<td>transsulfuration</td>
</tr>
<tr>
<td>Hcy</td>
<td>homocysteine</td>
</tr>
<tr>
<td>Cth</td>
<td>cystathionine</td>
</tr>
<tr>
<td>CGL</td>
<td>cystathionine-$\gamma$-lyase</td>
</tr>
</tbody>
</table>
REFERENCES


