De novo biosynthesis of azide by a promiscuous *N*-nitrosylase

Antonio Del Rio Flores[‡#], Rui Zhai[‡#], David W. Kastner[^+], Kaushik Seshadri[‡], Siyue Yang[♦], Kyle De Matias[‡], Yuanbo Shen[♦], Wenlong Cai[‡], Maanasa Narayanamoorthy[♦], Nicholas B. Do[‡], Zhaoqiang Xue[‡], Dunya Al Marzooqi[‡], Heather J. Kulik[+*] and Wenjun Zhang[‡*]

‡ Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, California 94720 United States. ♦ Department of Chemistry, University of California Berkeley, Berkeley, California 94720 United States. ^ Department of Bioengineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 United States. + Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 United States.

# Equal contribution

*email: hjkulik@mit.edu, wjzhang@berkeley.edu

Abstract

Azides are energy-rich compounds endowed with diverse representation in broad scientific disciplines, including material science, synthetic chemistry, pharmaceutical science, and chemical biology. Despite ubiquitous usage of the azido group, the underlying biosynthetic pathways for its formation remain largely unknown. Here we report the characterization of what is to our knowledge the first biosynthetic pathway for de novo azide construction. We demonstrate that Tri17, a promiscuous ATP and nitrite-dependent enzyme, catalyzes organic azide synthesis through sequential *N*-nitrosation and dehydration of aryl hydrazines. Through biochemical, structural, and computational analyses, we further propose a plausible molecular mechanism for azide biosynthesis that sets the stage for future biocatalytic applications and biosynthetic pathway engineering.

Introduction

The azide moiety is an electron-rich functionality with extensive applications in drug discovery, synthetic chemistry, material science, and chemical biology. For example, azidothymidine (AZT), a synthetic analog of thymidine, is a well-known medication used in the management and treatment of HIV-1 by inhibiting nucleoside reverse transcriptases[1,2]. The π bond in the azido group can be easily polarized to promote strong exothermic dissociation reactions to release reactive nitrene groups and molecular nitrogen, with applications requiring high-energy-density-materials in propellants and energetic polymers[3,4]. Organic azides are also common synthetic precursors for triazole and tetrazole-containing compounds[1,5]. Notably, the azide-alkyne

cycloaddition reaction has revolutionized the field of chemical biology due to its “bio-orthogonal” and “clickable” features. It enabled selective imaging and studies of azide/alkyne-labelled molecules, ranging from small-molecule metabolites and drugs to macromolecules such as glycans, lipids, proteins, and nucleic acids[6–10].

The broad application of azide in chemistry has raised significant interest in the azidation reaction using enzymes. Specifically, multiple non-heme iron halogenases and oxygenases have been engineered to catalyze azide transfer, typically from an inorganic azido donor to a few selected substrates[11–17]. Alternatively, the azide group might be directly biosynthesized in situ without feeding the bio-orthogonal moiety itself. This new approach would enable the biogenesis of azide under mild, physiological conditions, allowing for tunable production of this bio-orthogonal handle for in vivo labeling with low metabolic background. However, to date, there is no enzyme known to promote de novo azide biosynthesis.

Our recent work on the biosynthesis of triacsins, a family of natural products possessing a conserved *N*-hydroxytriazene moiety, revealed two enzymes responsible for the two consecutive N-N bond forming steps[18]. Tri28 was shown to be a glycine-utilizing, hydrazine-forming enzyme consisting of a cupin and a methionyl-tRNA synthetase-like domain[18]. This didomain enzyme and its homologs have recently emerged to be a new family of “hydrazine synthetases” with promising biocatalytic potential[18–23]. The second N-N bond forming enzyme in triacsin biosynthesis, Tri17, was shown to be an ATP and nitrite-dependent *N*-nitrosylase that completes *N*-hydroxytriazene installation on a linear alkyl hydrazone substrate (**Supplementary Fig. 1a**)[18]. Tri17 appeared to be promiscuous toward acyl chain modifications, but its biocatalytic potential remained elusive. Interestingly, several Tri17 homologs, such as CreM[24], Aha11[25] and AvaA6[26], were recently shown to also utilize nitrite to install a diazo moiety on different aryl amine substrates in the biosynthesis of cremeomycin, tasikamides, and avenalumic acid, respectively (**Supplementary Fig. 1**). These diazo biosynthetic pathways were proposed to proceed through similar *N*-nitrosylated intermediates, which subsequently undergo dehydration, although the exact mechanism and role of the enzymes have not been thoroughly investigated.

Here we probed the substrate specificity of Tri17 via biochemical and kinetic analyses. We revealed Tri17 to be highly promiscuous, promoting *N*-nitrosating reactions on a wide array of substrates, such as alkyl hydrazone, hydrazine, pyrrolidine, piperidine, aryl amine and aryl hydrazine. Notably, Tri17 converted multiple aryl hydrazine substrates to azido products, representing the first natural enzyme capable of de novo azide biosynthesis. Further structural and computational analyses provided additional insight into the catalytic mechanism of Tri17, including the enzyme-facilitated dehydration during azide formation.

Results

Tri17 is a promiscuous *N*-nitrosylase that recognizes substrates beyond linear alkyl hydrazones. Our prior biosynthetic study of triacsins suggested that Tri17 converts a terminal hydrazone moiety of a fully unsaturated dodecenoic acid to an *N*-hydroxytriazene moiety (**Supplementary Fig. 1a**)[18]. Due to the instability of this proposed substrate[27], we reconstituted the activity of Tri17 by utilizing a surrogate substrate (**1**) which demonstrated a k_{cat}/K_m of 135 ± 0.7 $mM^{-1}min^{-1}$ (**Fig. 1**)[18]. Although Tri17 appeared to be promiscuous towards acyl chain modifications, we found that it was unable to recognize 12-aminododecanoic acid nor 2-hydrazineylideneacetic acid[18]. To better understand the substrate scope of Tri17, we initially tested a pool of ~20 compounds varying in alkyl chain length, modification, and the nitrogen-containing moiety (**Fig. 1, Supplementary Fig. 2-6, Supplementary Note 1**). All reactions were monitored by liquid chromatography high-resolution mass spectrometry (LC-HRMS) analysis. We employed both targeted *N*-nitrosylated species detection as well as an untargeted product search via comparative metabolomics to assess the substrate recognition of Tri17. We observed that in addition to hydrazone, Tri17 recognized hydrazine and nitrogen-containing heterocycles, such as pyrrolidine and piperidine, to generate *N*-nitrosylated products, albeit with a decreased catalytic efficiency (**Fig. 1, Supplementary Fig. 3-5**). Although Tri17 did not recognize butylhydrazine (C4), the enzyme was tolerant towards alkyl chains of medium chain length (~C6-C11) with a preference for longer chain length, especially for hydrazone substrates. The enzyme was also not sensitive to the methylation of alkyl chains at various positions. Tri17 did not recognize a few other tested compounds containing hydrazide, urea, pyridine, etc. (**Supplementary Fig. 6**).

Considering that enzymes homologous to Tri17 recognized various aryl amines to catalyze diazo formation[24–26], we next probed the possible recognition of aryl amines by Tri17. One of the Tri17 homologs, Aha11 (39% sequence similarity) was also purified and assayed for comparison with Tri17 (**Supplementary Fig. 2**). The catalytic activity of Aha11 was biochemically reconstituted using a surrogate substrate (**12**) as previously reported (**Fig. 2**)[25]. Interestingly, Tri17 recognized **12** and produced the same diazo product (**13**) with a catalytic efficiency comparable to that of Aha11 (**Fig. 2, Supplementary Fig. 7**). Altogether, Tri17 demonstrated to be a promiscuous *N*-nitrosylase that can recognize diverse substrates, including an aryl amine that is drastically different from an alkyl hydrazone, the proposed native substrate of Tri17[18].

Tri17 orchestrates azidation of aryl hydrazine substrates. Since Tri17 demonstrated relaxed substrate specificity and catalyzed diazo formation from an aryl amine substrate, presumably via an *N*-nitrosylated intermediate followed by dehydration, it raised an interesting possibility that Tri17 may also promote azido formation on an aryl hydrazine substrate via a similar mechanism. We first selected compound **14** to probe this potential activity of Tri17 as both **14** and the predicted azido product **15** were commercially available. The biochemical assay containing **14**, ATP, nitrite, and Tri17 led to the production of two new products which were identified to be 1-azidonaphthalene (**15**) and 1-naphthylamine (**16**), respectively, by comparison with the authentic standards based on the LC-HRMS analyses and ultraviolet-visible (UV-vis) spectra (**Fig. 3, Supplementary Fig. 8 and 9**). Neither product was observed from negative controls in which Tri17, ATP, nitrite, or **14** was omitted. The utilization of ^{15}N-nitrite also led to the expected mass spectral shift of **15**, further confirming the utilization of nitrite for azide formation by Tri17 (**Fig. 3**). **16** was most likely a degraded product of **15** upon spontaneous loss of N_2, which was also observed in the pure standard solution of **15**. Tri17 exhibited a good catalytic efficiency towards **14** with a k_{cat}/K_m of 49.5 ± 5.1 $mM^{-1}min^{-1}$ by monitoring substrate consumption.

After demonstrating that Tri17 could promote azide formation, we next probed additional aryl hydrazine compounds to expand its substrate scope in generating organic azides. When hydralazine (**17**), a vasodilator drug[28–30] was used as a substrate, two new products, **18** and **19**, were detected based on comparative metabolomics analyses using LC-HRMS (**Fig. 4a,**

Supplementary Fig. 10-11). While **18** was proposed to be an *N*-nitrosylated species based on the HRMS and activity of Tri17, **19** was assigned to be the expected azide product according to the HRMS and UV spectrum, as well as the detection of a click reaction product between 5-hexyn-1-ol and **19** (**Supplementary Fig. 12**). A time-course analysis of this reaction showed the transient nature of **18** and a steady increase in the production of **19,** suggesting that **18** may be an intermediate en route to azide formation (**Fig. 4b**). The kinetic parameters of Tri17 towards **17** were further determined through quantification of substrate consumption (k_{cat}= 30.1 ± 3.7 min^{-1}, K_M= 0.31 ± 0.06 mM, k_{cat}/K_M= 97.1 ± 5.7 $mM^{-1}min^{-1}$), which are comparable to the best recognized substate **1** (**Fig. 4c**). Although **19** was produced under acidic conditions (pH<5) without Tri17 (**Supplementary Fig. 11e**), the enzyme was required to produce **18** and **19** around physiological pH (**Fig. 4a)**. We further sought to enzymatically synthesize 7-azido-4-methyl-coumarin (**22**), which is a fluorogenic probe for H_2S detection widely used in vitro, in living cells, and cardiac tissues[31,32]. A biochemical assay containing **20**, Tri17, ATP, and nitrite resulted in the production of two new products, the presumed *N*-nitrosylated intermediate (**21)** and the expected azido product (**22**), which was confirmed by comparing it to an authentic standard (**Extended Data Fig. 1, Supplementary Fig. 13**). Although the catalytic efficiency of Tri17 towards **20** (k_{cat}/K_M= 22.8 ± 1.1 $mM^{-1}min^{-1}$) has yet to be improved, these results suggested that Tri17 is a promiscuous enzyme towards various aryl hydrazines with biocatalytic potential to produce azides under physiological conditions. It is notable that we were unable to detect corresponding azido or *N*-nitrosylated products from biochemical assays with aryl hydrazines and Aha11, further highlighting the unique biocatalytic potential of Tri17 (**Fig. 3-4, Extended Data Fig. 1**).

Tri17 promotes azido natural product biosynthesis from dihydralazine. The azido group is rare in nature, appearing in only one known natural product: 6-azidotetrazolo[5,1-a]phthalazine (ATPH, **25**) from *Karenia brevis*, a toxin-producing dinoflagellate associated with red tides in the Gulf of Mexico[33–35]. While the biosynthesis of ATPH remains obscure, we reasoned that double azidation of dihydralazine (**23**) by Tri17 may yield a short-lived diazide intermediate that undergoes intramolecular azide–tetrazole isomerization to yield ATPH (**Fig. 5**). Such isomerization reactions have been reported in the chemical synthesis of ATPH and are common amongst

azido functionalities linked to a carbon atom that is adjacent to an annular nitrogen[33,36,37]. The biochemical assay of Tri17 and **23** produced a major product (**24**) that was consistent with a single azidation step (**Fig. 5a, Supplementary Fig. 14**). Interestingly, Tri17 successfully recognized **24** and catalyzed the formation of **25** which was confirmed to be ATPH by comparison to a synthetic standard (**Fig. 5b, Supplementary Note 2, Supplementary Fig. 15**). A copper-free click reaction using dibenzocyclooctyne-PEG4-Fluor 545 with the Tri17 reaction product (**25**) yielded **26**, further indicating the presence of one azide moiety in **25** which is consistent with literature that a second isomerization of **25** to give bis(tetrazole) does not take place (**Fig. 5c, Supplementary Fig. 16-17**)[33].

Tri17 facilitates dehydration of an *N*-nitrosylated intermediate to form azide. Our biochemical assays of Tri17 with **17** or **20** suggested the intermediacy of an *N*-nitrosylated species preceding synthesis of the azido synthon via dehydration (**Fig. 4, Extended Data Fig. 1**). However, it remained unclear whether the dehydration step was spontaneous in the reaction buffer or facilitated by Tri17. We found that the HPLC-purified **19** remained stable over a 6-hour period in the assay buffer (**Supplementary Fig. 11)**. On the other hand, the *N*-nitrosylated intermediate, **18**, was unstable with a half-life of ~30 min in our assay buffer, but it did not readily form **19** and its degradation product remained obscure (**Supplementary Fig. 10d**). Considering the instability of **18**, we obtained a mixture of **18** and **19** by setting up the Tri17 reaction with **17** for a short duration (30 min), followed by organic solvent extraction of the reaction mixture. The amount of **19** in this mixture remained unchanged over a 6-hour period while the amount of **18** decreased, which was consistent with the stability results of both compounds and suggested that **18** did not undergo spontaneous dehydration to form **19** (**Fig. 4d, Supplementary Fig. 10d-e)**. We next used this mixture of **18** and **19** as substrates for subsequent Tri17 assays without addition of ATP nor nitrite. The production of **19** was observed in a time-dependent manner with the addition of Tri17, demonstrating the role of Tri17 in enzyme-mediated dehydration of **18** (**Fig. 4d**). The addition of ATP and Mg^{2+} to this assay did not significantly enhance the activity of Tri17 in this dehydration reaction (**Extended Data Fig. 2)**.

To better understand the dehydration of the *N*-nitrosylated intermediate, we also performed computational modeling to investigate the likelihood of

spontaneous dehydration of the *N*-nitrosylated intermediate in an aqueous environment. We reasoned that after the Tri17-catalyzed *N*-nitrosation of **17**, the resulting intermediate (**18P**) may undergo tautomerization to generate *cis* or *trans* isomers (**18C** and **18**), respectively. Azide formation would presumably conclude through dehydration of either **18C** or **18** **(Supplementary Fig. 18**). We first examined the energy profiles and geometric characteristics of the proposed intermediates—**18P**, **18**, **18C**—involved in the tautomerization reaction and the subsequent conversion to **19** in solvent. Our calculations revealed a sequential decrease in energy from the initial intermediate **18P** to either **18** or **18C**, and then to the final product **19** (**Supplementary Fig. 18, Table 4**). Specifically, **19** was significantly more stable than **18P**, showing a relative energy difference of -39.3 kcal/mol as evaluated at the local coupled cluster level of theory (see Methods).

Next, we investigated the tautomerization of **18P** to either the *trans* (**18**) or *cis* (**18C**) conformer. The triacsin family of natural products were reported in the *trans* conformation, suggesting that this conformation was favored for the Tri17 products[18,38–40]. However, notable differences in the structure of **18P** compared to the triacsins may influence whether the *trans* or *cis* product is favored. Our computational results showed a modest activation barrier for the tautomerization of **18P** to **18** of 6 kcal/mol using the local coupled cluster level of theory (**Supplementary Table 5, Supplementary Fig. 19**). The tautomerization of **18P** to **18C** had a higher activation barrier of 8.34 kcal/mol (**Supplementary Table 5, Supplementary Fig. 20**). These results thus showed that the conversion of **18P** to either **18** or **18C** can occur spontaneously due to the low energy barriers for tautomerization with the *trans* configuration being more energetically favorable.

Subsequently, we investigated the putative spontaneous conversion of **18** or **18C** to **19** in an aqueous environment. The calculated barriers for the initiation of the conversion from **18** to **19** were prohibitively high for both proposed mechanisms (**Supplementary Fig. 21**). Our calculations revealed the reaction to be endothermic by 28.4 kcal/mol for the deprotonation of the amine of the *N*-hydroxytriazene moiety and endothermic by 53.1 kcal/mol for the protonation of the *N*-hydroxy group (**Supplementary Fig. 22-23**). For the *cis* conformer **18C**, the barrier to **19** via an intramolecular proton transfer to the *N*-hydroxy group calculated at the local coupled-cluster level of theory, was slightly lower at 23.8 kcal/mol (**Supplementary Table 6,**

Supplementary Fig. 24). Nonetheless, the overall significant magnitude of the energetic barriers for the spontaneous dehydration of the *N*-nitrosylated intermediate to form azide strongly pointed to the necessity of enzymatic catalysis for the efficient production of **19,** consistent with our biochemical assay results.

Probing catalytic mechanisms of Tri17 by structural and computational analyses. To provide structural details for the azido-forming reaction catalyzed by Tri17, we solved the X-ray crystal structure of Tri17 at 2.4 Å resolution in its apo form and generated structural models docked with the various substrates and intermediates (**Fig. 6**). The overall structure of Tri17 adopts the characteristic two-domain fold for the adenylate-forming enzyme superfamily (ANL superfamily)[41,42], consisting of the large *N*-terminal (light blue) and the small *C*-terminal domain (pale green), with the active site located at the two domain interfaces. A search for a structural homolog of Tri17 by the Dali Program[43] gave the closest functionally characterized protein as 4-coumarate CoA ligase (4-CLs)[44](PDB: 3TSY, *Z* score = 36.4, root mean square deviation (RMSD) of 3.6 Å for 391 C_α atoms, and 20% sequence identity). Despite screening of co-crystallization and soaking conditions, our efforts to obtain a diffraction-quality crystal structure of Tri17 complexed with ATP/AMP cofactors has not been successful. Nevertheless, the high structural similarity of Tri17 with 4-CLs made it possible to investigate the ATP/AMP binding sites through structural alignment. A crystal structure of 4-CL isoform 2 (Nt4CL2)[41] in complex with Mg^{2+} and ATP (PDB: 5BSM) was chosen to be aligned with the apo structure of Tri17, with a RMSD of 2.69 Å for C_α atoms of the overall sequence (**Supplemental Fig. 25**). The conserved residues involved in the recognition of ATP were identified via structural comparison and confirmed through mutagenesis studies. Substitution of any of these residues (S299, F338, E342, I363, D421, H433, R436, K539) to alanine greatly reduced or abolished the *N*-nitrosylation of compound **1** (**Fig. 6a-b**).

The ANL superfamily is well known to adopt a large conformational change between the adenylation and nucleophilic reactions[41]. The small *C*-terminal domain typically rotates to an extent of 140° to switch between the conformation for the adenylation reaction (Con_{Ad}) and the conformation for the following nucleophilic reaction (Con_{Nuc}). The current apo structure was crystallized in its adenylation conformation. To provide insights into the

substrate promiscuity of Tri17, a structural model representing Con_{Nuc} was generated[45] and nitroso-AMP was first docked into the corresponding pocket for the ATP/AMP cofactor using Autodock Vina[46,47]. To validate the docking model accuracy, alignment of the cofactor surrounded by conserved residues was conducted between the docking pose and the AMP from Nt4CL2 in its Con_{Nuc} (PDB: 5BSR), yielding an RMSD of 0.18 Å (**Supplementary Fig. 26**). Subsequent examination of this docked model revealed a potential binding pocket[48] constructed by residues from both the *N*- and *C*-terminal domains (**Fig. 6c**). The size of the pocket was calculated[49] to be 420 $Å^3$, which is much larger than the volume size of the substrate **1** (200 $Å^3$). To investigate the binding mode for different substrates, **1**, **2**, **6** and **17** were chosen as representative substrates for linear hydrazone, linear hydrazine, heterocyclic amine, and aryl hydrazine, respectively, and were docked into the binding pocket via Autodock Vina (**Extended Data Fig. 3, Supplementary Fig. 27**)[46,47]. The docking results indicated a substrate binding mode where the nucleophile reaches nitroso-AMP through a narrowed "neck" formed by H230, F273 and G446 (**Fig. 6c**, **Extended Data Fig. 3**). The narrowed tunnel would likely hold both the nitroso group and nucleophile in a proper position for the nucleophilic reaction to happen. The importance of these residues was verified through the mutagenesis experiments, in which Tri17_H230A and Tri17_F273A dramatically reduced *N*-nitrosation of **1**, while Tri17_G446S abolished the reaction by blocking the tunnel (**Fig. 6b**). The large binding pocket most likely offers ample space to accommodate substrates of varying sizes, explaining the broad substrate specificity observed for Tri17.

To further shed light on the dehydration step that leads to the formation of an azido product, docking simulations were performed to place **18** and **2P** into the putative substrate binding pocket (**Extended Data Fig. 4**). A receptor structure was first generated by docking AMP into the current model and the docking accuracy was validated in a similar manner as described above (RMSD = 0.41 Å). **18** and **2P** were subsequently docked into the receptor structure and the poses with the highest binding affinity were chosen for subsequent analyses via comparison with the binding mode of the corresponding substrate. The results showed that **2P** reached out more extensively toward AMP due to the flexibility of the alkyl chain, while the bulkier phthalazine moiety in **18** tended to be accommodated in the larger

portion of the pocket, placing the N-N-N-O moiety toward H229, a key residue that may facilitate the protonation/deprotonation in the dehydration process (**Fig. 6c**, **Extended Data Fig. 4**). The importance of H229 was tested by conducting a time-course analysis of the Tri17_H229F reaction with **17**. We observed the transient nature of **18** similar to the time course experiment with the wild-type Tri17 (**Extended Data Fig. 5, Fig. 4b**); however, no **19** was formed with Tri17_H229F. Consistently, Tri17_H229F completely lost the ability to catalyze the dehydration of **18** to **19** (**Fig. 4d**). These observations provided compelling evidence regarding the unique catalytic activity of Tri17, which not only exhibited broad substrate promiscuity to accept aryl hydrazine substrates to produce azide, but also assumed a catalytic role in the dehydration process of azido group formation.

Discussion

Our in-depth study centering on the substrate promiscuity of Tri17 allowed us to uncover its innate ability to recognize a broad range of substrates, which is consistent with the predicted large substrate binding pocket. Notably, aryl hydrazines were among the best substrates to be recognized. Tri17 was capable of nearly matching the activity of Aha11 in recognizing **12** for diazo formation, but Aha11 was unable to recognize any tested aryl hydrazine substrates (**Fig. 2-5, Extended Data Fig. 1**). Given the similarity and difference in substrate specificity between Tri17 and Aha11, it prompted us to conduct bioinformatic analyses to probe the diversity of ATP-utilizing *N*-nitrosylases. Our sequence similarity network (SSN) analyses showed that Aha11 and CreM clustered together in one group, while Tri17 belonged to a distinct group (**Extended Data Fig. 6a**). Consistently, a phylogenetic tree of Tri17 showed that it was in a different clade from its known homologs (**Extended Data Fig. 6b**). Interestingly, our SSN analyses suggested that two other groups (groups 3 and 4) of Tri17 homologs may contribute to additional functional diversity within the ATP-utilizing *N*-nitrosylase space (**Extended Data Fig. 6a**). We further hypothesize that mining natural product biosynthetic gene clusters encoding a Tri17 homolog may be fruitful for discovering new azide biosynthetic enzymes and new azido-containing natural products. These metabolites may have escaped discovery due to instability or intermediacy which was observed for the intermediate of 3-diazoavenalumic acid in avenalumic acid biosynthesis[26].

One of the most intriguing reaction products synthesized by Tri17 was the purported natural product ATPH (**25**) produced by *Karenia brevis*. ATPH

possesses unique structural features not found in any other N-N containing natural product, such as an equal number of carbon and nitrogen atoms, and six N-N bonds that make up an azido group, tetrazolo ring, and phthalazine ring[34,50]. Although the biosynthetic pathway of ATPH in *Karenia brevis* is yet to be elucidated, this work leads to an interesting hypothesis that this natural product could be generated by a Tri17 homolog from dihydralazine (**23**), a vasodilator drug that has long been used for the treatment of hypertension[28–30]. The biosynthesis of **25** catalyzed by Tri17 further uncovered a new biosynthetic route to tetrazole ring formation. Tetrazole-containing small molecules are well known for their potent biological activities in medicinal chemistry as antifungal, antibacterial, antihypertensive, antitumor, and anticancer agents[51–53]. We thus have indicated an additional facet to the biocatalytic potential of Tri17 beyond azido formation.

While Tri17 recognized a wide array of substrates to yield *N*-nitrosylated products, different fates for hydrazine substrates were observed considering the subsequent dehydration reaction to form azide. Our biochemical, structural, and computational analyses strongly indicated that Tri17 played a catalytic role in dehydration and further suggested a molecular mechanism for different product outcomes. Specifically, the shape of the substrate binding pocket of Tri17 is characteristic of a "gourd", where the *N*-nitrosylated alkyl product extends more inside toward AMP, while the bulkier aryl product such as **18** would fall back due to the steric hindrance, with phthalazine accommodated within the large part of the binding pocket formed with aromatic residues. This specific orientation of **18** would place the terminal N-N-N-O moiety close to the hydrophilic residues within the oxyanion hole (N110, E445 and H229), where H229 forms direct hydrogen bonding with the N1 and O4, functions as the catalytic residue for the protonation and deprotonation process, and leads to the formation of azide via dehydration (**Fig. 6b-c, Extended Data Fig. 7**, **Supplementary Fig. 28**). It's noteworthy that H229 may also play an important role in defining the shape of the binding pocket, as the mutagenesis assay of Tri17_H229F with **1** resulted in a 47% relative activity compared to the wild-type Tri17, while the H229I variant abrogated *N*-nitrosylation of **1** (**Fig. 6b**). Besides H229, other residue(s) in the active site may also aid in the dehydration of **18**. Further insights into substrate binding and catalysis will be facilitated with a substrate bound structure of Tri17, which is currently in progress.

In conclusion, we investigated the biocatalytic potential of Tri17 and uncovered its innate ability to *N*-nitrosylate various types of substrates, including hydrazones, hydrazines, aryl amines, and nitrogen-containing

heterocycles. Tri17 is the first identified enzyme capable of de novo biosynthesis of the azido synthon, enabling the biogenesis of azide under mild, physiological conditions. Additional biochemical, structural, and computational analyses shed light on the catalytic mechanism of Tri17, particularly in promoting dehydration of the *N*-nitrosylated intermediate to form azide. This work paves the way for further mechanistic interrogation and biocatalytic applications of Tri17 and its homologs for azide and other unique functional group biosynthesis.

Acknowledgements

We would like to thank Dr. Jeff Pelton for helping us with NMR spectroscopic analyses. This research was financially supported by a grant to W.Z. from the NIH (R01GM136758). A.D.R.F. was financially supported by the Blavatnik Innovation Fellowship and UC Berkeley Chancellor's Fellowship. X-ray data was collected at Beamline 8.3.1 of the Advanced Light Source, a DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231, which is supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. This work was supported in part by the National Science Foundation (CBET-1704266 and CBET-1846426 for H.J.K and D.W.K.). H.J.K. holds a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund, which supported this work (H.J.K. and D.W.K.). The authors would like to thank Douglas Millar for assistance with bioinformatic analyses.

Author contributions

A.D.R.F. designed the experiments, performed biochemical and bioinformatic analysis of Tri17 and variants, aided with structural analysis of Tri17, analyzed the data, and wrote the manuscript. R.Z. designed the experiments, performed structural and modeling analysis for Tri17, analyzed the data, and wrote the manuscript. D.W.K. designed computational experiments, performed calculations, analyzed the data, and contributed to the writing of the computational work. K.S. helped conduct in vitro experiments, chemical synthesis, protein purification and kinetic characterization of substrates. W.C. analyzed the NMR data and aided with chemical synthesis. S.Y., Y.S., K.D.M., M.N. and N.D. aided in protein

purification, construction of plasmids, and repeating biochemical assays for this study. N.B.D. aided in protein purification, biochemical assays, construction of plasmids, and interpretation of LC-MS data. Z.X. aided R.Z. with the structural work of Tri17. D.A.M helped collect and analyze the NMR data. H.J.K. designed computational experiments, analyzed the data, and wrote the manuscript. W.Z. designed the experiments, analyzed the data, and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Correspondence and requests for materials should be addressed to H.J.K and W.Z.

References

1. Lin, T.-S. & Prusoff, W. H. Synthesis and biological activity of several amino analogs of thymidine. *J. Med. Chem.* **21**, 109–112 (1978).
2. Liu, Y. *et al.* Anti-HIV agent azidothymidine decreases Tet(X)-mediated bacterial resistance to tigecycline in Escherichia coli. *Commun. Biol.* **3**, 1–10 (2020).
3. Huynh, M. H. V., Hiskey, M. A., Chavez, D. E., Naud, D. L. & Gilardi, R. D. Synthesis, characterization, and energetic properties of diazido heteroaromatic high-nitrogen C-N compound. *J. Am. Chem. Soc.* **127**, 12537–12543 (2005).
4. Agrawal, J. P. & Hodgson, R. D. *Organic Chemistry of Explosives*. *Organic Chemistry of Explosives* (2007). doi:10.1002/9780470059364
5. Bräse, S., Gil, C., Knepper, K. & Zimmermann, V. Organic azides: An exploding diversity of a unique class of compounds. *Angew. Chemie - Int. Ed.* **44**, 5188–5240 (2005).
6. Thirumurugan, P., Matosiuk, D. & Jozwiak, K. Click chemistry for drug development and diverse chemical-biology applications. *Chem. Rev.* **113**, 4905–4979 (2013).
7. Zhu, X., Liu, J. & Zhang, W. De novo biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11**, 115–120 (2015).
8. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48**, 6974–6998 (2009).
9. Grammel, M. & Hang, H. C. Chemical reporters for biological discovery. *Nat. Chem. Biol.* **9**, 475–484 (2013).
10. Del Rio Flores, A. *et al.* Biosynthesis of Isonitrile- and Alkyne-Containing Natural Products. *Annu. Rev. Chem. Biomol. Eng.* **13**, 1–24 (2022).

11. Matthews, M. L. *et al.* Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014).

12. Kim, C. Y. *et al.* The chloroalkaloid (−)-acutumine is biosynthesized via a Fe(II)- and 2-oxoglutarate-dependent halogenase in Menispermaceae plants. *Nat. Commun.* **11**, 1–7 (2020).

13. Neugebauer, M. E. *et al.* A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019).

14. Voss, M., Honda Malca, S. & Buller, R. Exploring the Biocatalytic Potential of Fe/α-Ketoglutarate-Dependent Halogenases. *Chem. - A Eur. J.* **26**, 7336–7345 (2020).

15. Gomez, C. A., Mondal, D., Du, Q., Chan, N. & Lewis, J. C. Directed Evolution of an Iron(II)- and α-Ketoglutarate-Dependent Dioxygenase for Site-Selective Azidation of Unactivated Aliphatic C−H Bonds**. *Angew. Chemie* **202301370**, (2023).

16. Rui, J. *et al.* Directed evolution of non-heme iron enzymes to access a non-natural radical-relay C(sp3)−H azidation. *Science (80-.).* **376**, 869–874 (2022).

17. Chan, N. H. *et al.* Non-Native Anionic Ligand Binding and Reactivity in Engineered Variants of the Fe(II)- and α-Ketoglutarate-Dependent Oxygenase, SadA. *Inorg. Chem.* **61**, 14477–14485 (2022).

18. Del Rio Flores, A. *et al.* Biosynthesis of triacsin featuring an N-hydroxytriazene pharmacophore. *Nat. Chem. Biol.* **17**, 1305–1313 (2021).

19. Matsuda, K. *et al.* Discovery of Unprecedented Hydrazine-Forming Machinery in Bacteria. *J. Am. Chem. Soc.* **140**, 9083–9086 (2018).

20. Zhao, G. *et al.* Molecular basis of enzymatic nitrogen-nitrogen formation by a family of zinc-binding cupin enzymes. *Nat. Commun.* **12**, 1–10 (2021).

21. He, H. Y., Niikura, H., Du, Y. L. & Ryan, K. S. Synthetic and biosynthetic routes to nitrogen-nitrogen bonds. *Chem. Soc. Rev.* 2991–3046 (2022).

22. Matsuda, K. *et al.* A Natural Dihydropyridazinone Scaffold Generated from a Unique Substrate for a Hydrazine-Forming Enzyme. *J. Am. Chem. Soc.* **144**, 12954–12960 (2022).

23. Kuga Arima, Satoko Akiyama, Kazuo Shin-ya, K. & Matsuda, and T. W. Carrier Protein Mediated Formation of the Dihydropyridazinone Ring in Actinopyridazinone Biosynthesis. *Angew. Chemie Int. Ed.* (2023). doi:10.1002/anie.202305155

24. Waldman, A. J. & Balskus, E. P. Discovery of a Diazo-Forming Enzyme in Cremeomycin Biosynthesis. *J. Org. Chem.* **83**, 7539–7546 (2018).

25. Ma, G. L. *et al.* Biosynthesis of Tasikamides via Pathway Coupling and Diazonium-Mediated Hydrazone Formation. *J. Am. Chem. Soc.* **144**, 1622–1633

(2022).

26. Kawai, S., Hagihara, R., Shin-ya, K., Katsuyama, Y. & Ohnishi, Y. Bacterial Avenalumic Acid Biosynthesis Includes Substitution of an Aromatic Amino Group for Hydride by Nitrous Acid Dependent Diazotization. *Angew. Chemie* **134**, (2022).

27. Yoshida, Keizo; Okamoto, Masanori ;Umehara, Kazuyoshi; Iwami, Morita; Kohsaka, Masanobu;Aoki, Hatsuo;Imanaka, H. Studies on New Vasodilators, WS-1228 A and B I. Discovery, Taxonomy, Isolation and Characterization. *J. Antibiot. (Tokyo).* **XXXV**, 157–163 (1981).

28. Reece, P. A. Hydralazine and related compounds: Chemistry, metabolism, and mode of action. *Med. Res. Rev.* **1**, 73–96 (1981).

29. Arce, C. *et al.* Hydralazine target: From blood vessels to the epigenome. *J. Transl. Med.* **4**, 1–16 (2006).

30. De Flora, S. *et al.* In vivo and in vitro genotoxicity of three antihypertensive hydrazine derivatives (hydralazine, dihydralazine, and endralazine). *Environ. Mutagen.* **4**, 605–619 (1982).

31. Thorson, M. K., Majtan, T., Kraus, J. P. & Barrios, A. M. Identification of cystathionine β-synthase inhibitors using a hydrogen sulfide selective probe. *Angew. Chemie - Int. Ed.* **52**, 4641–4644 (2013).

32. Chen, B. *et al.* Fluorescent probe for highly selective and sensitive detection of hydrogen sulfide in living cells and cardiac tissues. *Analyst* **138**, 946–951 (2013).

33. Nunez Avila, A. G. *et al.* Surprising Chemistry of 6-Azidotetrazolo[5,1-a]phthalazine: What a Purported Natural Product Reveals about the Polymorphism of Explosives. *J. Org. Chem.* **87**, 6680–6694 (2022).

34. Blair, L. M. & Sperry, J. Natural products containing a nitrogen-nitrogen bond. *J. Nat. Prod.* **76**, 794–812 (2013).

35. Hossain, M. B., van der Helm, D., Sanduja, R. & Alam, M. Structure of 6-azidotetrazolo[5,1-a]phthalazine, C8H4N8, isolated from the toxic dinoflaggelate Gymnodinium breve. *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* **41**, 1199–1202 (1985).

36. Tišler, M. Some Aspects of Azido-Tetrazolo Isomerization. *Synthesis (Stuttg).* **03**, 123–136 (1973).

37. Krivopalov, V. P., Baram, S. . G., Denisov, A. Y. & Mamatyuk, V. I. Azide-tetrazole tautomerism of diazidodiazines and their benzo analogs. *Bull. Acad. Sci. USSR, Div. Chem. Sci.* **38**, 1839–1844 (1989).

38. Omura, Satoshi, Tomoda, Hiroshi, Xu Min, Qin, Takahashi, Yoko, Iwai, Y. Triacsins, New Inhibitors of Acyl-CoA Synthetase Produced by Streptomyces SP. 1211–1218 (1986).

39. Twigg, F. F. *et al.* Identifying the Biosynthetic Gene Cluster for Triacsins with an N-Hydroxytriazene Moiety. *ChemBioChem* **20**, 1145–1149 (2019).

40. Hiroshi, T., Kazuaki, I. & Satoshi, O. Inhibition of acyl-CoA synthetase by triacsins. *Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab.* **921**, 595–598 (1987).

41. Li, Z. & Nair, S. K. Structural Basis for Specificity and Flexibility in a Plant 4-Coumarate:CoA Ligase. *Structure* **23**, 2032–2042 (2015).

42. Gulick, A. M. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827 (2009).

43. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).

44. Wang, Y., Yi, H., Wang, M., Yu, O. & Jez, J. M. Structural and kinetic analysis of the unnatural fusion protein 4-coumaroyl-CoA ligase::stilbene synthase. *J. Am. Chem. Soc.* **133**, 20684–20687 (2011).

45. Yang, Z. *et al.* UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J. Struct. Biol.* **179**, 269–278 (2012).

46. Oleg Trott, A. J. O. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **31**, 455–461 (2009).

47. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).

48. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).

49. Smith, R. H. B., Dar, A. C. & Schlessinger, A. PyVOL: a PyMOL plugin for visualization, comparison, and volume calculation of drug-binding sites. *bioRxiv* **c**, 816702 (2019).

50. Waldman, A. J., Ng, T. L., Wang, P. & Balskus, E. P. Heteroatom-Heteroatom Bond Formation in Natural Product Biosynthesis. *Chem. Rev.* **117**, 5784–5863 (2017).

51. Myznikov, L. V., Vorona, S. V. & Zevatskii, Y. E. Biologically active compounds and drugs in the tetrazole series. *Chem. Heterocycl. Compd.* **57**, 224–233 (2021).

52. Zou, Y., Liu, L., Liu, J. & Liu, G. Bioisosteres in drug discovery: Focus on tetrazole. *Future Med. Chem.* **12**, 91–93 (2020).

53. Ostrovskii, V. A., Trifonov, R. E. & Popova, E. A. Medicinal chemistry of tetrazoles. *Russ. Chem. Bull.* **61**, 768–780 (2012).

Online Methods

Materials. Phusion High-Fidelity PCR Master Mix (Thermo Scientific) was used for PCR reactions. Restriction enzymes were purchased from Thermo Scientific. All chemicals used in this work were obtained from Alfa Aesar, Enamine, Sigma-Aldrich or Fisher Scientific, unless otherwise noted. ^{15}N-sodium nitrite (purity >98%) and NMR solvents were purchased from Cambridge Isotope Laboratories, Inc.

Bacterial Strains and Growth Conditions. *Escherichia coli* strains were cultivated on lysogeny broth (LB) agar plates or liquid terrific broth (TB). Growth media was supplied with 50 μg/mL of kanamycin.

Construction of Plasmids for Expression in *E. coli*. One vector was used for *E. coli* induced expression of recombinant proteins using polyhistidine tags. The plasmid pET-24b(+) was used for the expression of Tri17 and Aha11 with *C*-terminal polyhistidine tags. The *Nde*I and *Xho*I sites were used for the restriction digest and ligation-based installation of Tri17. The *Nde*I and *HindIII* sites were used for the restriction digest and ligation-based installation of Aha11. Tri17 was amplified from *Streptomyces tsukubaensis NRRL 18488* as listed in **Table S2**. The *aha11* gene was purchased as a gBlock from Integrated DNA technologies after performing codon optimization suitable for E. *coli* expression.

Tri17 variants were constructed by PCR using the Agilent QuikChange II Site-Directed Mutagenesis kit. pET24b(+)-Tri17 was used as a template. Reactions were conducted according to the manufacturer's protocol. The PCR program began at 95 °C for 30 s, followed by 18 cycles of 68 °C for 1 min, 68 °C for 7 min, and final extension at 68 °C for 7 min. The template DNA was digested with 10 units of DpnI for 1 h at 37 °C, and the remaining PCR product was transformed into *Escherichia coli* XL-1 Blue competent cells by heat-shock. The introduction of mutation(s) was confirmed with DNA sequencing. Oligonucleotides utilized in this study were purchased from Integrated DNA Technologies. All oligonucleotides and strains used in this study are listed in **Table S1**.

Expression and Purification of Recombinant Proteins. The expression and purification for all proteins used in this study followed the same general procedure for polyhistidine tag purification as detailed here. Expression strains were grown at 37 °C in 1 L of TB in a shake flask supplemented with 50 μg/mL of kanamycin to an OD_{600} of 0.6 at 250 rpm. The shake flask was then placed over ice for 10 min and induced with 250 μM of isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were then incubated for 20 hours at 16 °C at 250 rpm to undergo protein expression. Subsequently, the cells were harvested by centrifugation (6,371 x g, 15 min, 4 °C), and the supernatant was removed. The cell pellet was resuspended in 30 mL of lysis buffer (25 mM HEPES pH 8.0, 500 mM NaCl, 5 mM imidazole) and cells were lysed by sonication on ice. Cellular debris was removed by centrifugation (27,216 x g, 1 hour, 4 °C) and the supernatant was filtered with a 0.45 μm filter before batch binding. Ni-NTA resin (Qiagen) was added to the filtrate at 2 mL/L of cell culture, and the samples were allowed to nutate for 1 hour at 4 °C. The protein-resin mixture was loaded onto a gravity flow column. The flow through was discarded and the column was then washed with approximately 25 mL of wash buffer (25 mM HEPES pH 8.0, 100 mM NaCl, 20 mM imidazole) and tagged protein was eluted in approximately 15 mL of elution buffer (25 mM HEPES pH 8.0, 100 mM NaCl, 250 mM

imidazole). The whole process was monitored with a Bradford assay. Purified proteins were concentrated and exchanged into exchange buffer (25 mM HEPES pH 8.0, 100 mM NaCl) using Amicon ultra filtration units. After two rounds of buffer exchange and concentration, the purified enzyme was removed, and glycerol was added to a final concentration of 10% (vol/vol). Enzymes were subsequently flash-frozen in liquid nitrogen and stored at -80 ºC. The presence and purity of purified enzymes was assessed using SDS-PAGE and the concentration was determined using a NanoDrop UV-vis spectrophotometer (Thermo Fisher Scientific).

To prepare the pure protein for the crystal screening, the concentrated eluate from the Ni-NTA column was filtered and subjected to ion-exchange chromatography using the MonoQ column (MonoQ™ 4.6/100 PE, Cytiva). The protein was eluted with a linear gradient of 50–1000 mM NaCl in 50 mM Tris pH 8.0. The Tri17 protein was further purified on a Superdex 200 pg column (Cytiva) via gel filtration chromatography. The target protein was eluted with 20 mM Tris pH 8.0 containing 20 mM NaCl, and concentrated to 15 mg/ml with an Amicon Ultra-4 filter at 4 °C. The presence and purity of the enzyme was monitored by SDS-PAGE.

The approximate molecular weight and yield for each protein are the following: Tri17 (61.5 kDa, 34 mg/L) and Aha11 (57.0 kDa, 5 mg/L). The Tri17 variants had purification yields that were nearly the same compared to the wild-type enzyme (±5%).

Tri17 Activity Assays. Reactions were performed at room temperature for 30 min (unless otherwise noted) in 100 µL of 50 mM Tris pH 8.0, 0.5 mM primary substrate, 5 mM sodium nitrite (or ^{15}N-sodium nitrite), 5 mM ATP, 5 mM $MgCl_2$, and 20 µM Tri17. Primary substrates were dissolved in DMSO to ensure full solubility and assays were maintained at a final concentration of 2% DMSO (vol/vol). After the incubation period, the reaction was quenched with two volumes of chilled methanol. The precipitated protein was removed by centrifugation (15,000 x g, 10 min) and the supernatant was used for analysis. LC-HRMS analysis was performed using an Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. At least three independent replicates were performed for each assay, and representative results are shown.

Comparative Metabolomics. Tri17 assays utilizing various primary substrates were analyzed via LC-HRMS using an Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. Peak picking and comparative metabolomics were performed using MSDial with peak lists exported to Microsoft Excel.

Determination of Tri17 Kinetic Parameters Towards Primary Substrates. Assays were performed in triplicate in 50 µL of 50 mM Tris pH 8.0 containing 5 mM nitrite, 5 mM ATP, 5 mM $MgCl_2$, and 20 µM Tri17. The concentration of primary substrate was varied depending on the kinetic activity (see individual Michaelis Menten graphs). The incubation times for the reactions were 1 min, 5 min, 10 min, 20 min, and 40 min, which were used to determine the initial velocity of the reaction. After each incubation period, the reactions were quenched with two volumes of chilled methanol. The precipitated protein was

removed by centrifugation (15,000 x g, 5 min) and the supernatant was used for analysis. LC-HRMS analysis was performed using an Agilent technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. Product concentration was estimated by constructing a standard curve using an authentic triacsin A standard for *N*-nitrosylated products (**5P-11P**)[18]. Substrate consumption of **14**, **17**, and **20** was monitored to estimate the product concentration for kinetic analyses. Kinetic parameters were determined and plotted using GraphPad Prism 9.

Click Reaction between 19 and 5-hexyn-1-ol. A 200 µL Tri17 activity assay using hydralazine (**17**) as a substrate was conducted as detailed in a previous section and quenched with 400 µL of cold methanol. The precipitated protein was removed by centrifugation (15,000 x g, 10 min). In a 1.7-mL Eppendorf tube, 450 µL of the reaction supernatant was mixed with 50 µL of 5 mM 5-hexyn-1-ol. A premixed solution containing 10 µL of 20 mM $CuSO_4$ and 15 µL of 50 mM (THPTA) was sequentially added. 25 µL of 100 mM sodium ascorbate was added to start the reaction. The reaction mixture was allowed to incubate overnight at room temperature. The click reaction product was extracted from the click reaction mixture with 1:1 (v/v) ethyl acetate (3x). The organic phase was dried under nitrogen and resuspended in 100 µL of MeOH for LC-HRMS analyses. LC-HRMS analysis was performed using an Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. At least three independent replicates were performed for each assay, and representative results are shown.

Copper-free Click Reaction with 25 and dibenzocyclooctyne-PEG4-Fluor 545. A 200 µL Tri17 activity assay using **23** as a substrate was conducted as detailed in a previous section and quenched with 400 µL of cold methanol and the precipitated protein was removed by centrifugation (15,000 x g, 10 min). In a 1.7-mL Eppendorf tube, 450 µL of the reaction supernatant was mixed with 150 µL of 2 mM dibenzocyclooctyne-PEG4-Fluor 545 (dissolved in DMSO). The reaction mixture was allowed to incubate overnight at room temperature. The click reaction product was extracted from the click reaction mixture with 1:1 (v/v) ethyl acetate (3x). The organic phase was dried under nitrogen and resuspended in 100 µL of MeOH for LC-HRMS analyses. LC-HRMS analysis was performed using an Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. At least three independent replicates were performed for each assay, and representative results are shown.

Analysis of supernatant from Tri17 assay with 17. A large-scale assay was performed in triplicate in 350 µL of 50 mM Tris pH 8.0 containing 5 mM nitrite, 5 mM ATP, 5 mM $MgCl_2$, 1 mM **17**, and 20 µM Tri17. After a 30 min incubation period, the solution was spin-filtered using a 2-kDa Amicon spin filter to remove protein residues and isolate the flowthrough. The flowthrough was analyzed at the following time points: 10 min, 30 min, 1 hr, 2 hr, 3 hr, and 6 hr. Immediately before LC-MS analysis, 50 µL of the flowthrough was diluted with 100 µL of cold methanol. LC-HRMS analysis was performed using an

Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. The amounts of **18** and **19** were estimated from developing a calibration curve of **17**.

Biochemical assays with Tri17 supernatant. The flowthrough from the previous section was extracted three times with equal volume ethyl acetate. The organic fractions were dried under nitrogen and resuspended in 350 µL of 50 mM Tris pH 8.0 containing 20 µM Tri17. The assay mixture was analyzed at the following time points: 10 min, 30 min, 1 hr, 2 hr, and 3 hr. Immediately before LC-MS analysis, 50 µL of the assay was diluted with 100 µL of cold methanol. The precipitated protein was removed by centrifugation (15,000 x g, 5 min) and the supernatant was used for analysis.LC-HRMS analysis was performed using an Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized. The amount of **19** was estimated from developing a calibration curve of **17**.

Isolation of 18 and 19 from Tri17 biochemical assay and stability analysis. A 500 µL biochemical assay with Tri17 and **17** was conducted as outlined in a previous section. After a 30-minute incubation period, the enzymatic assay was quenched with 1 mL of cold methanol. The precipitated protein was removed by centrifugation (15,000 x g, 5 min) and the supernatant was used for further HPLC purification. HPLC purification was carried out using an Agilent 1200 systems and a Grace Alltima C18 column (150 x 10 mm) with a 5-95% acetonitrile gradient at a flow rate 2.5 mL/min by monitoring UV_{280nm}. All HPLC fractions were screened utilizing LC-HRMS and fractions containing **18** and **19** were pooled together and dried under nitrogen, respectively. **18** and **19** were dissolved in 50 mM Tris pH 8.0 and 50 µL aliquots were analyzed over the following time points: 10 min, 30 min, 1 hr, 2 hr, 3 hr, and 6 hr. LC-UV-HRMS analysis was performed using an Agilent Technologies 6545 Q-TOF LC-MS equipped with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A water/acetonitrile mobile phase with 0.1% (vol/vol) formic acid with a linear gradient of 2-98% acetonitrile at a flow rate of 0.5 mL/min was utilized.

QM Calculations. We performed all quantum mechanical (QM) calculations using ORCA version 5.0.3[54]. The reactant, product, and transition state (TS) geometries were initially optimized using density functional theory (DFT) with the global hybrid B3LYP functional, together with a 6-31G* basis set[55–57]. We applied a semi-empirical DFT-D3 dispersion correction with default Becke-Johnson damping[58] to all DFT calculations. An implicit conductor-like polarizable continuum model (C-PCM) with a dielectric constant (ε) of 80 was also employed to approximate a water environment[59,60].

To identify candidate transition states, we performed constrained geometry scans at the same level of theory by incrementally fixing the reaction coordinate while optimizing all other degrees of freedom. The reactant, product, and the maximum energy structure from the scan were subsequently used as inputs for a zoomed climbing image (CI) Nudged elastic band (NEB) calculation as implemented in ORCA 5.0.3.[61]. The zoomed-NEB calculations consisted of an initial NEB with 50 frames followed by a second NEB with an

additional 50 frames performed around the identified CI using the default ORCA cutoff criteria. The geometry and hessian of the CI identified with zoomed-NEB was then employed as an initial guess for a partitioned rational-function optimization (P-RFO) calculation[62]. For the final optimized TS from P-RFO, we performed a frequency calculation to confirm the presence of a single imaginary frequency corresponding to the transition state.

We next optimized the geometries of the reactant and product with B3LYP and the quadruple-ζ def2-QZVPP basis set with additional polarization functions. We also ran P-RFO calculations at the same level of theory for the transition state. Single point energy calculations were then performed on the reactants, products and transition states with B3LYP/def2-TZVPP and def2-QZVPP basis sets and the energies were extrapolated to the complete basis set (CBS) limit using the two-point formula[54,63,64] as implemented in ORCA. We then performed single points on the optimized geometries at the DLPNO-CCSD(T)/def2-TZVPP and def2-QZVPP levels of theory and extrapolated to the CBS limit with the two-point formula. Given that C-PCM is not implemented for DLPNO-CCSD(T) in ORCA[65], we computed a solvent correction by evaluating the Møller–Plesset second-order perturbation theory (MP2)[66] interaction energy difference between the solvated and gas-phase states with the following equation:

$$E_{CCSD(T)\ Solvated} = E_{CCSD(T)\ Gas\ Phase} + (E_{MP2\ Solvated} - E_{MP2\ Gas\ Phase})$$

We performed the MP2 calculations in ORCA using the same basis set used for DLPNO-CCSD(T) with C-PCM and a dielectric constant of 80 for implicit solvent[67]. The coordinates of the B3LYP/def2-QZVPP geometry optimized reactants, products, and transition states are included in the Source Data zip file.

Crystallization and Structural Determination. The crystals of Tri17 were obtained at 16 ◦C in 100 mM Tris pH 8.0, containing 20% PEG3350, 100 mM KSCN, with 15 mg/ml of the purified Tri17 solution, by using the sitting-drop vapor-diffusion method. Crystals appear in two weeks. Crystals were transferred into the cryoprotectant solution (reservoir solution with 20% (v/v) ethylene glycol), and then flash cooled at −173 ◦C in a nitrogen-gas stream. We used BL8.3.1 at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory (LBNL) to collect X-ray diffraction data sets. A 1.00 Å beam wavelength was used for data collection. The diffraction data sets for Tri17 were processed and scaled using the XDS program package[68] and Aimless[69]. The initial phases of the Tri17 structure were determined by molecular replacement, using the homology model generated by alphaFold[70,71] as the search template. We performed molecular replacement with *Phaser* in the PHENIX software suite[72]. The initial phases were calculated with *AutoBuild* in PHENIX[73]. The structures were modified manually with Coot[74] and refined with *PHENIX_Refine*[75]. The final crystal data and intensity statistics are summarized in **Supplementary Table 7**. The Ramachandran statistics are as follows:

97.4% favored, 2.6% allowed for Tri17. All crystallographic figures were prepared with PyMOL (DeLano Scientific, http://www.pymol.org).

Structural Modelling and Docking Simulations. Homology model of Tri17 in its nucleophilic reaction conformation was generated with Modeller[45] using multiple tertiary structures in its Con_{Nuc} (PDB ID: 1PG4[76], 3E7W[77], 4G37[78], 5UPS[79],5BSR[41]) as templates. The generated model was relaxed to remove the potential clashes and geometry errors via energy minimization using steepest descent in GROMACS 2021 packages[80] with OPLS-AA/L all atom force field[81]. 800 minimization steps are performed with a maximum force convergence threshold of 1.0kJ/mol/nm. The model quality was subsequently examined by the Z-score calculated from ProsA Server[82], with a score of -10.01 within the range observed for the native set of proteins of the same size[83]. The Ramachandran plot, widely used to analyze the backbone conformation of protein structures, was calculated using MolProbity[84] to validate the main chain torsion angles to be stereo-chemically feasible. Outliers are checked individually to make sure they are not involved in the formation of the binding pocket (**Supplementary Fig. 29**). Docking simulations were performed with Autodock Vina[46,47], and ran at an exhaustiveness of 8. The grid box for AMP and nitroso-AMP was centered at X = 76.16 Y = 29.72, and Z = 10.55, with a grid box dimension of 13.90 Å × 10.00 Å × 10.60 Å, enclosing the conserved residues in the cofactor binding pocket. The grid box for other substrates was centered at X = 70.15 Y = 28.95, and Z = 18.38 with a grid box dimension of 19.88 Å × 20.14 Å × 19.43, including the residues constructing the potential binding pocket. The best poses with highest binding affinity were extracted and listed in **Supplementary Table 8**. To calculate the root means square deviation (RMSD), atoms in the docking ligand were first renumbered using PDBTools in PHENIX package, and then aligned using pair_fit command in PyMOL to fit the atom pairs between the docking ligand and PDB ligand (AMP in PDB ID 5BSR). A potential pocket was detected from CASTp[48] and the size of the pocket was calculated with PyVOL[49].

Construction of sequence similarity network of Tri17. The SSN network consisting of 1,471 Tri17 homologs was constructed by utilizing previously described methods with the EFI-Enzyme Similarity Tool web resource by using default parameters (https://efi.igb.illinois.edu/efi-est/tutorial_analysis.php)[85–87]. The SSN was visualized and analyzed using Cytoscape 3.10.0[88,89](https://cytoscape.org).

Construction of phylogenetic tree of Tri17. A total of 131 sequences were analyzed, 20 structural homologs from a Dali search[43] and 106 sequences retrieved from NCBI BLAST[90]. Sequences were aligned with the MUSCLE option[91] in MEGA-X Version 10[92] with default parameters. The bootstrap consensus tree inferred from 1000 replicates was used to represent the evolutionary history of the taxa for maximum-likelihood (ML) analysis.

Data availability. The authors declare that all data supporting the findings of this study are available within the paper, Supplementary Information, attached Zip file and the

Extended Data, and/or from the corresponding authors upon reasonable request. The coordinates and structure factor amplitudes for the apo structure of Tri17 was deposited to the Protein Data Bank (PDB) under accession code 8TF7.

54. Neese, F. Software update: The ORCA program system—Version 5.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **12**, 1–15 (2022).

55. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab Initio Calculation of Vibrational Absorption. *J. Phys. Chem.* **98**, 11623–11627 (1994).

56. Lee, C., Yang, W., and Parr, R. G. Development of the Collette-Salvetti correlation-energy formula into a function of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).

57. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **103**, 5648–5652 (1993).

58. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, (2010).

59. Lange, A. W. & Herbert, J. M. A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach. *J. Chem. Phys.* **133**, (2010).

60. York, D. M. A smooth solvation potential based on the conductor-like screening model. *J. Phys. Chem. A* **103**, 11040–11044 (1999).

61. Henkelman, G. & Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978–9985 (2000).

62. Baker, J. An algorithm for the location of transition states. *J. Comput. Chem.* **7**, 385–395 (1986).

63. Zhong, S., Barnes, E. C. & Petersson, G. A. Uniformly convergent n -tuple-augmented polarized (nZaP) basis sets for complete basis set extrapolations. I. Self-consistent field energies. *J. Chem. Phys.* **129**, (2008).

64. Helgaker, T., Klopper, W., Koch, H. & Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **106**, 9639–9646 (1997).

65. Riplinger, C. & Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **138**, (2013).

66. Møller, C. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46**, 618–622 (1934).

67. Klamt, A. & Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2* 799–805 (1993).

68. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

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

70. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

71. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

72. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

73. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

74. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

75. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).

76. Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M. & Escalante-Semerena, J. C. The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5′-propylphosphate and coenzyme A. *Biochemistry* **42**, 2866–2873 (2003).

77. Yonus, H. *et al.* Crystal structure of DltA: Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J. Biol. Chem.* **283**, 32484–32491 (2008).

78. Sundlov, J. A., Fontaine, D. M., Southworth, T. L., Branchini, B. R. & Gulick, A. M. Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism. *Biochemistry* **51**, 6493–6495 (2012).

79. Wang, N. *et al.* Natural separation of the acyl-CoA ligase reaction results in a non-adenylating enzyme article. *Nat. Chem. Biol.* **14**, 730–737 (2018).

80. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).

81. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **11**, 3499–3509 (2015).

82. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, 407–410 (2007).

83. Llorca, O. *et al.* The three-dimensional structure of an eukaryotic glutamine synthetase: Functional implications of its oligomeric structure. *J. Struct. Biol.* **156**,

469–479 (2006).

84. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

85. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169–4182 (2019).

86. Oberg, N., Zallot, R. & Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *J. Mol. Biol.* **435**, (2023).

87. Copp, J. N., Akiva, E., Babbitt, P. C. & Tokuriki, N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry* **57**, 4651–4662 (2018).

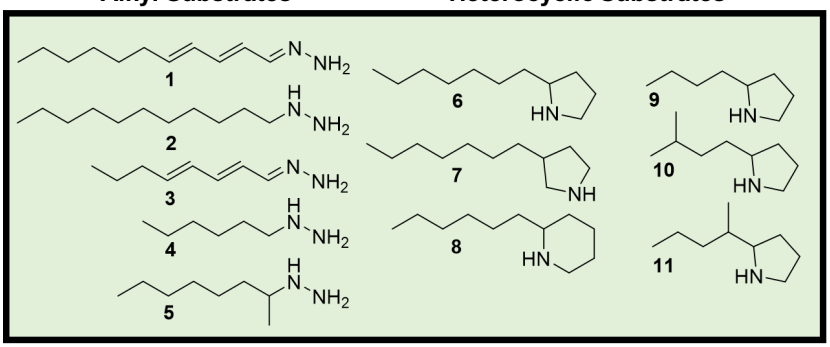
88. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **13**, 2498–2504 (2003).

89. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).

90. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

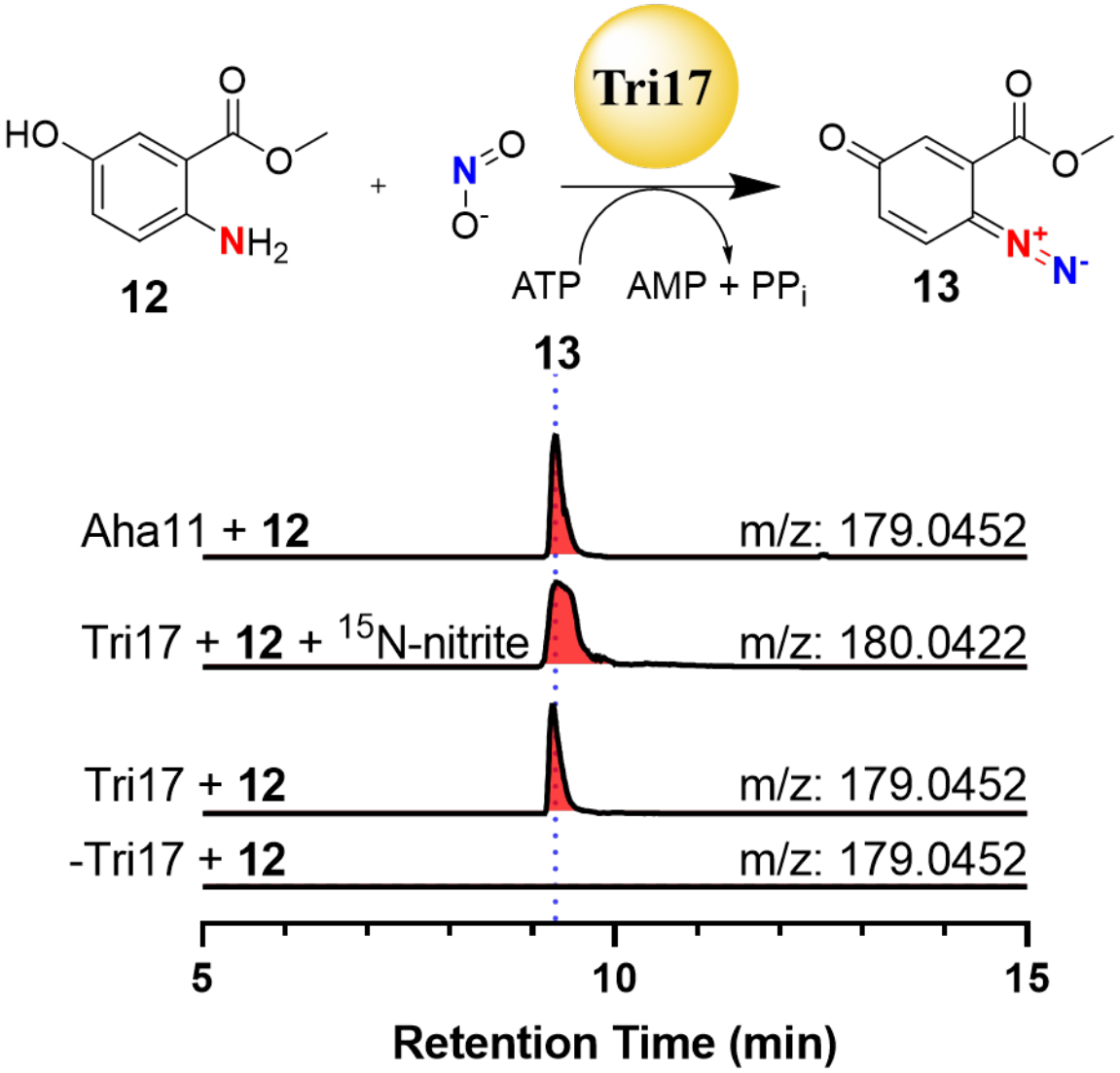
91. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

92. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).



Compound	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (mM^{-1}min^{-1})
1	40.6 ± 6.4	0.30 ± 0.05	135 ± 0.7
2	14.0 ± 0.8	1.1 ± 0.1	12.7 ± 0.4
3	34.3 ± 1.6	0.47 ± 0.03	73.0 ± 1.2
4	34.0 ± 2.7	0.84 ± 0.1	40.5 ± 2.7
5	ND	ND	47.2 ± 2.3
6	24.2 ± 1.6	1.3 ± 0.1	18.6 ± 0.2
7	15.0 ± 2.7	1.6 ± 0.3	9.4 ± 0.1
8	10.7 ± 0.9	3.6 ± 0.8	3.0 ± 0.4
9	ND	ND	14.3 ± 1.9
10	ND	ND	5.5 ± 0.2
11	ND	ND	4.7 ± 0.4

Fig. 1 Kinetic analysis of promiscuous Tri17. Tri17 kinetic parameters were determined for each substrate through LC-MS quantification. The parameters and uncertainty represent the average and standard deviation from three independently performed experiments, respectively. The kinetic parameters of Tri17 for **1** were previously determined[18]. ND: Not Determined.



951

952 **Fig. 2 Biochemical analyses of Tri17 and Aha11 with 12.** Extracted ion chromatograms
953 (EICs) demonstrating production of **13** from biochemical assays containing Tri17/Aha11,
954 ATP, nitrite, and **12.** Omission of any of these components resulted in abrogation of **13**.
955 Utilization of ^{15}N-nitrite resulted in the expected mass spectral shift. A 10-ppm mass error
956 tolerance was used for each trace. At least three independent replicates were performed
957 for each assay, and representative results are shown.

958

959

960

961

962

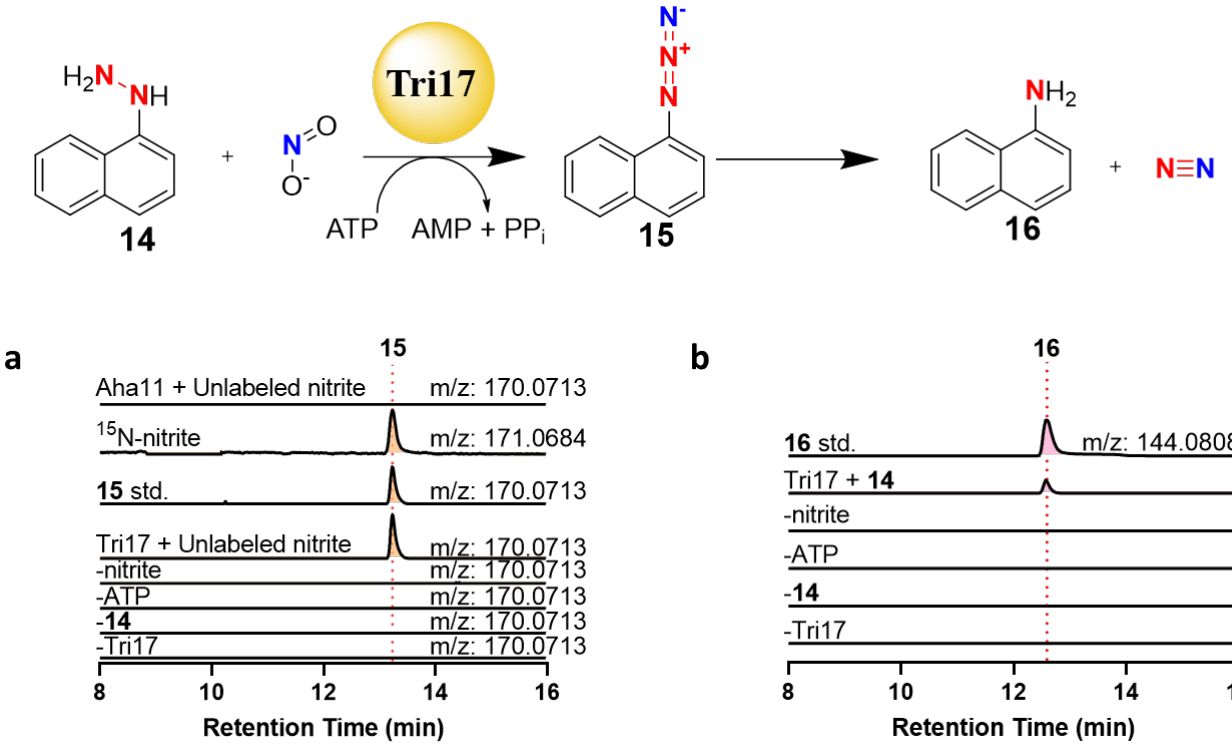


Fig. 3 Biochemical analysis of Tri17 with 14. A) EICs showing the production of **15** from assays containing Tri17, ATP, nitrite and **14**. Omission of any of these components resulted in abrogation of **15**. Utilization of ^{15}N-nitrite resulted in the expected mass spectral shift. No new products were detected when Aha11 was used in place of Tri17. B) EICs showing production of **16** as a degradation product of **15**. Authentic standards were utilized to confirm the production of **15** and **16**. A 10-ppm mass error tolerance was used for each trace. At least three independent replicates were performed for each assay, and representative results are shown.

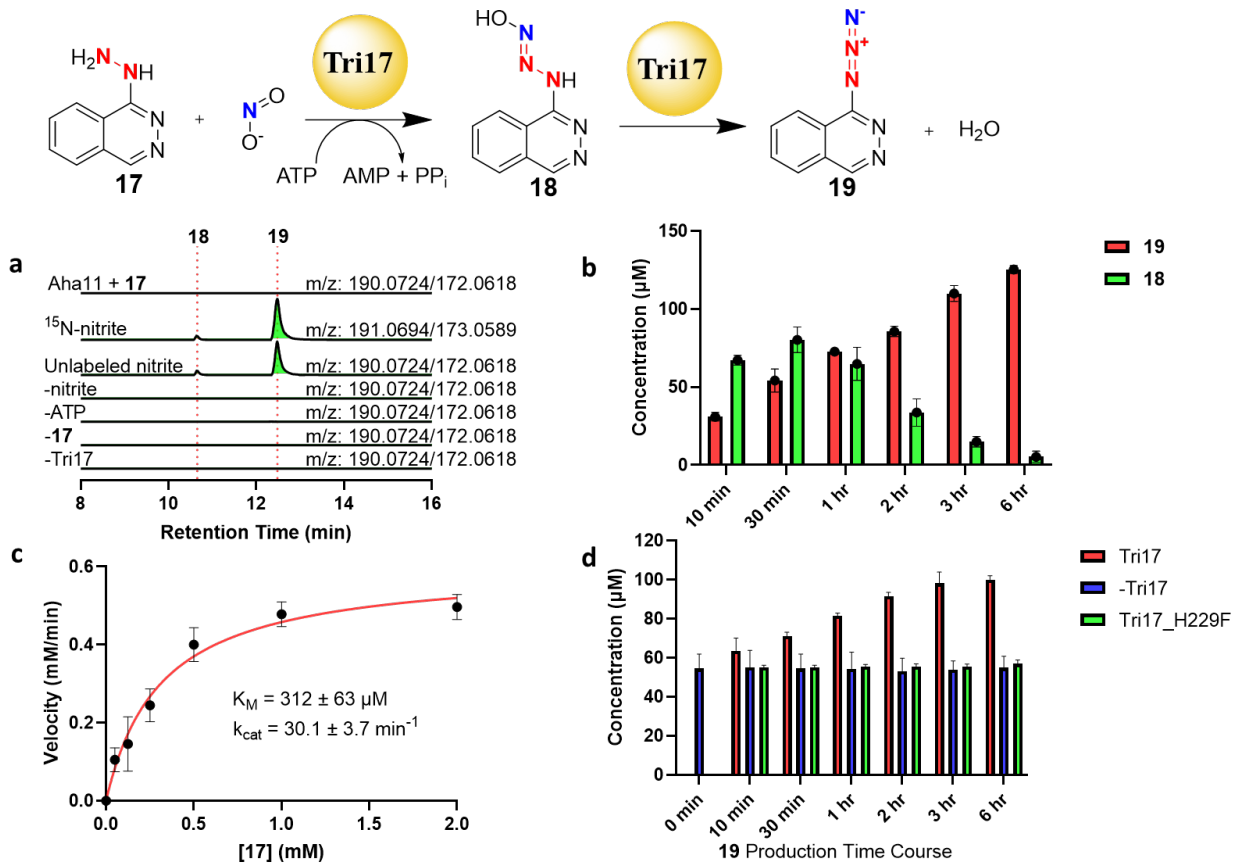


Fig. 4 Biochemical analysis of Tri17 with 17. A) EICs demonstrating the production of **18** and **19** from assays containing Tri17, ATP, nitrite and **17** after a 3-hour incubation period. Omission of any of these components resulted in abrogation of **18** and **19**. Utilization of ^{15}N-nitrite resulted in the expected mass spectral shift for both species. A 10-ppm mass error tolerance was used for each trace. No new products were detected when Aha11 was used in place of Tri17. b) Relative amounts of **18** and **19** quantified by LC-HRMS over a 6-hour time course of the Tri17 assay. Error bars correspond to standard deviation of the mean from three replicate experiments. c) Tri17 kinetic parameters towards **17**. The data points and error bars represent the average and standard deviation from three independent experiments, respectively. d) Analysis of **19** production from Tri17 assays. A Tri17 biochemical assay with **17** was first incubated at room temperature for 30 minutes and the protein was removed immediately using an Amicon spin filter (2 kDa MWCO). The reaction flowthrough was extracted with ethyl acetate, dried, and served as substrates (containing a mixture of **17**, **18**, and **19**) for new Tri17 reactions and the production of **19** was monitored in a time course. Tri17 wild-type and Tri17_H229F were used in new reactions together with no enzyme control. The data points and error bars represent the average and standard deviations from three independently performed experiments, respectively.

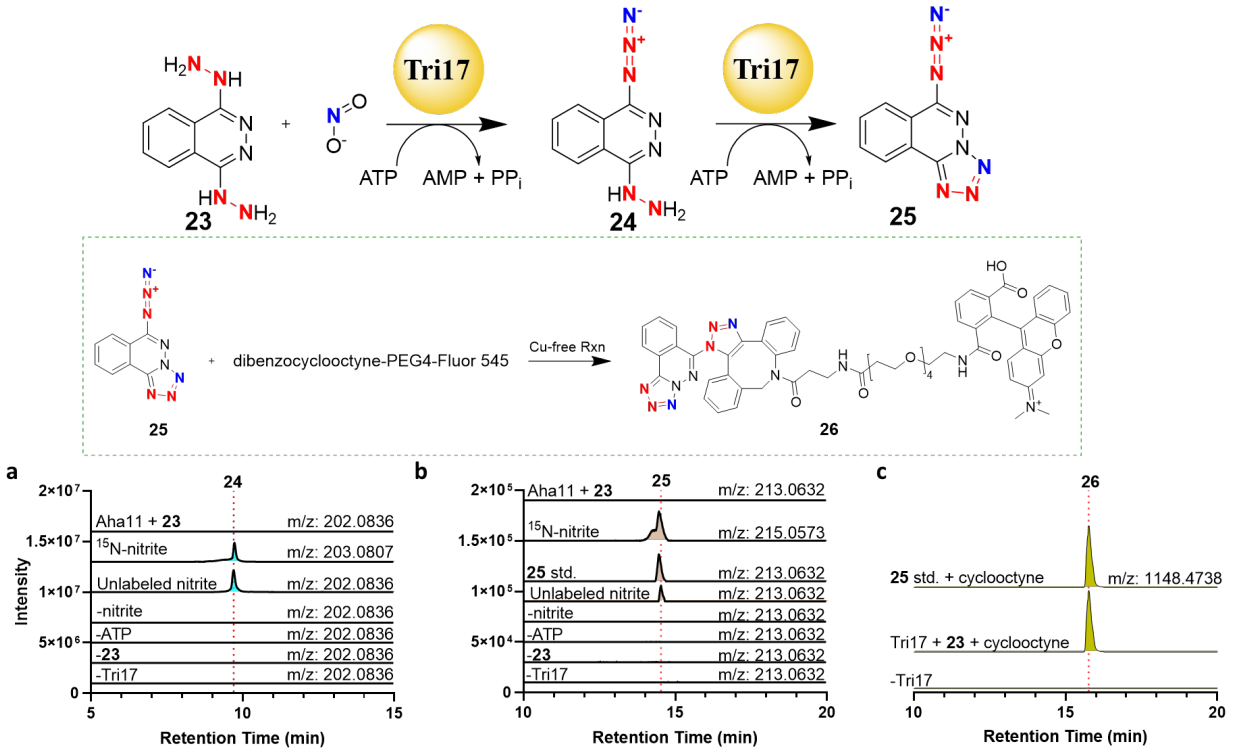
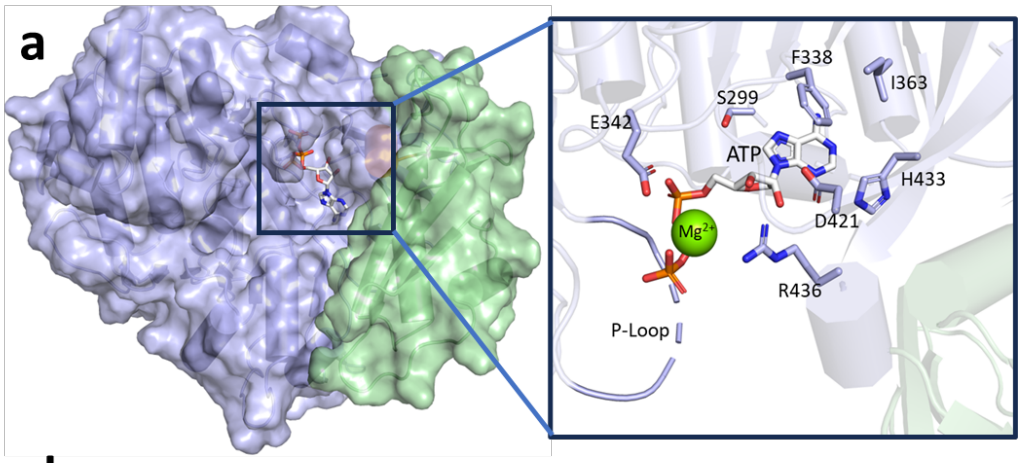
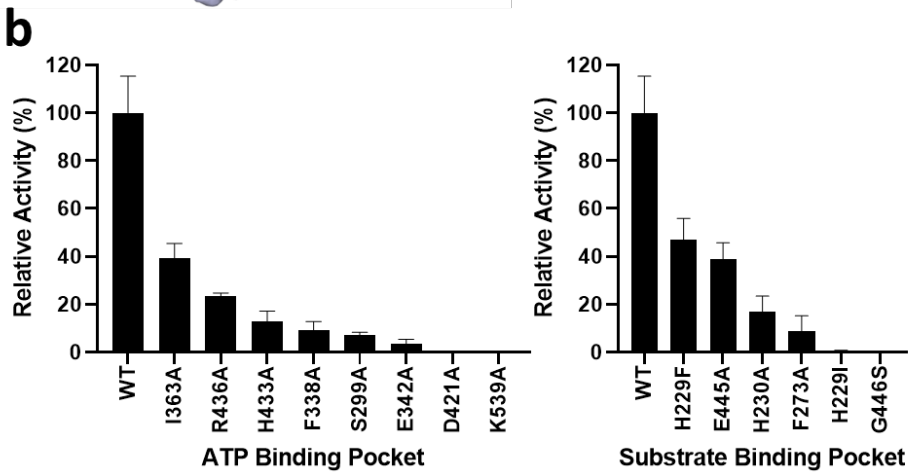
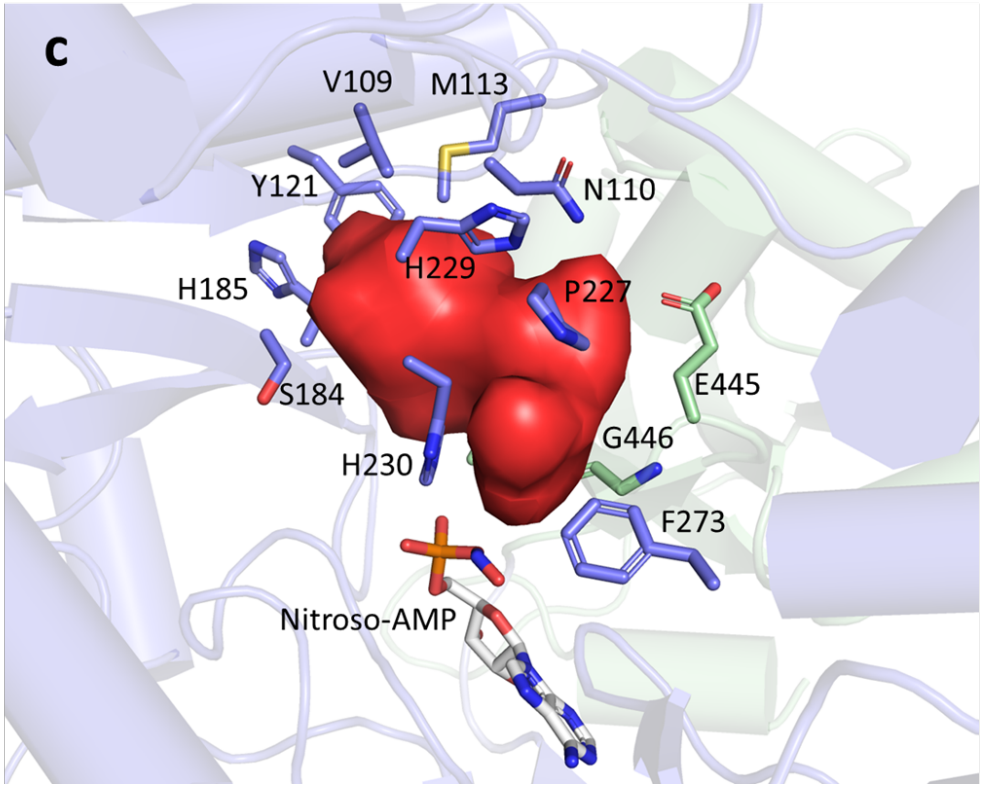


Fig. 5 Biochemical analysis of Tri17 with 23. a) EICs demonstrating the production of **24** from assays containing Tri17, ATP, nitrite and **23**. Omission of any of these components resulted in abrogation of **24**. Utilization of ^{15}N-nitrite resulted in the expected mass spectral shift. No new products were detected when Aha11 was used in place of Tri17. b) EICs demonstrating the production of **25** from assays containing Tri17, ATP, nitrite and **23**. Omission of any of these components resulted in abrogation of **25**. Utilization of ^{15}N-nitrite resulted in the expected mass spectral shift. **25** was synthesized and used to confirm the reaction product from the enzymatic assay. No new products were detected when Aha11 was used in place of Tri17. c) EICs demonstrating production of **26** from a copper-free click reaction between **25** and dibenzocyclooctyne-PEG4-Fluor 545. Omission of Tri17 resulted in abolishment of **26**. A 10-ppm mass error tolerance was used for each trace. At least three independent replicates were performed for each assay, and representative results are shown.

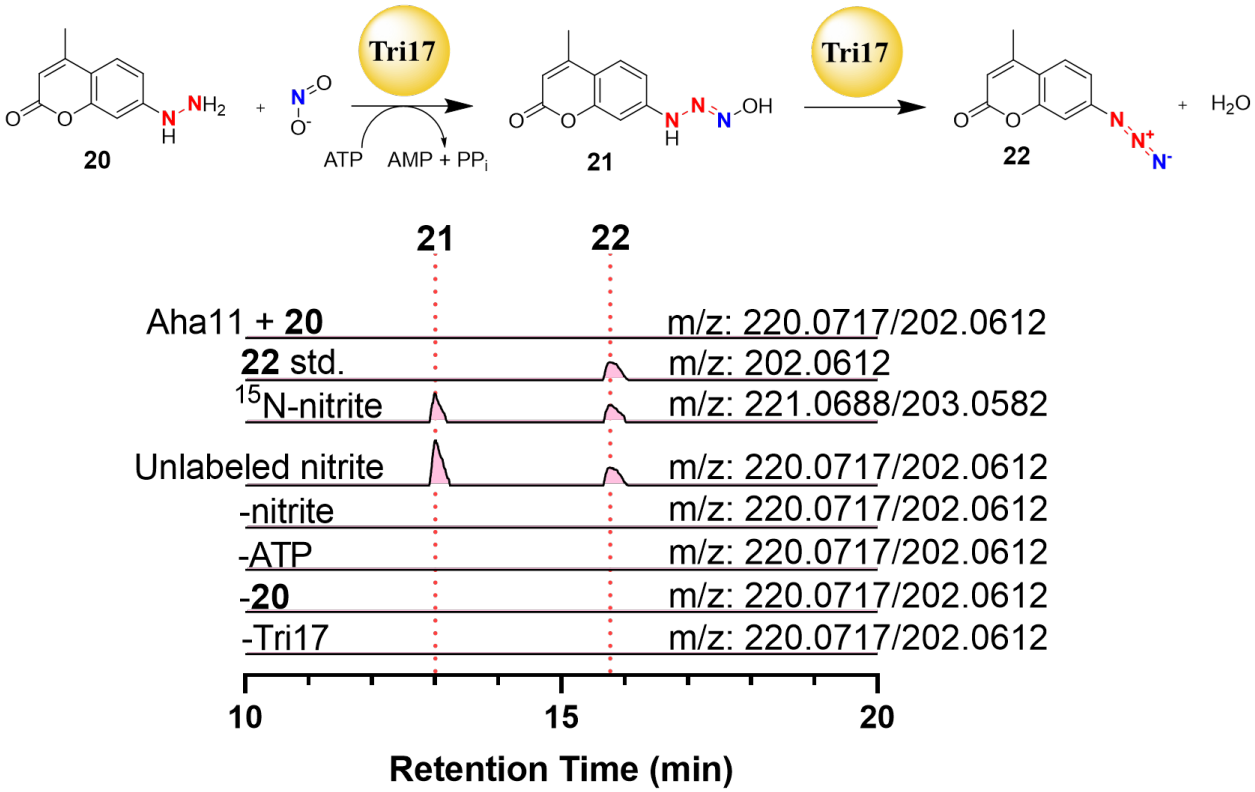

a
F338
I363
S299
E342
ATP
H433
D421
Mg2+
R436
P-Loop


b
Relative Activity (%)
120
100
80
60
40
20
0
WT
I363A
R436A
H433A
F338A
S299A
E342A
D421A
K539A
ATP Binding Pocket
Relative Activity (%)
120
100
80
60
40
20
0
WT
H229F
E445A
H230A
F273A
H229I
G446S
Substrate Binding Pocket

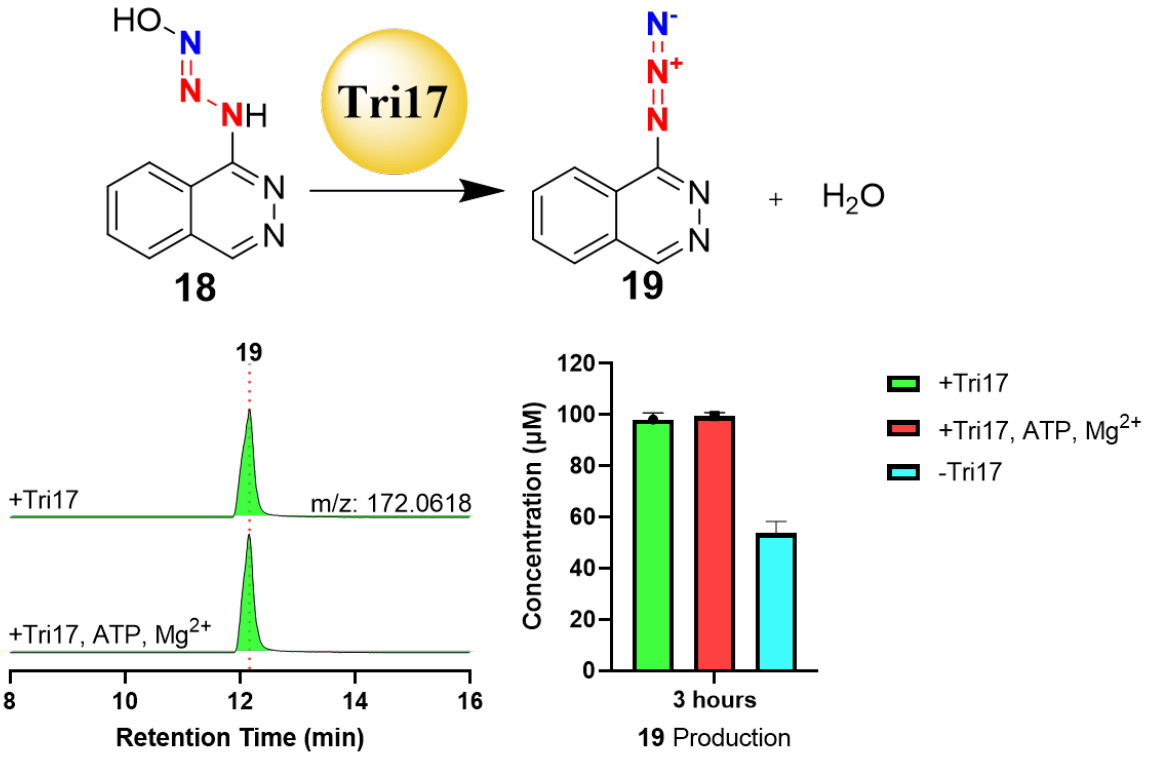

c
V109
M113
Y121
N110
H229
H185
P227
E445
S184
G446
H230
F273
Nitroso-AMP

1023

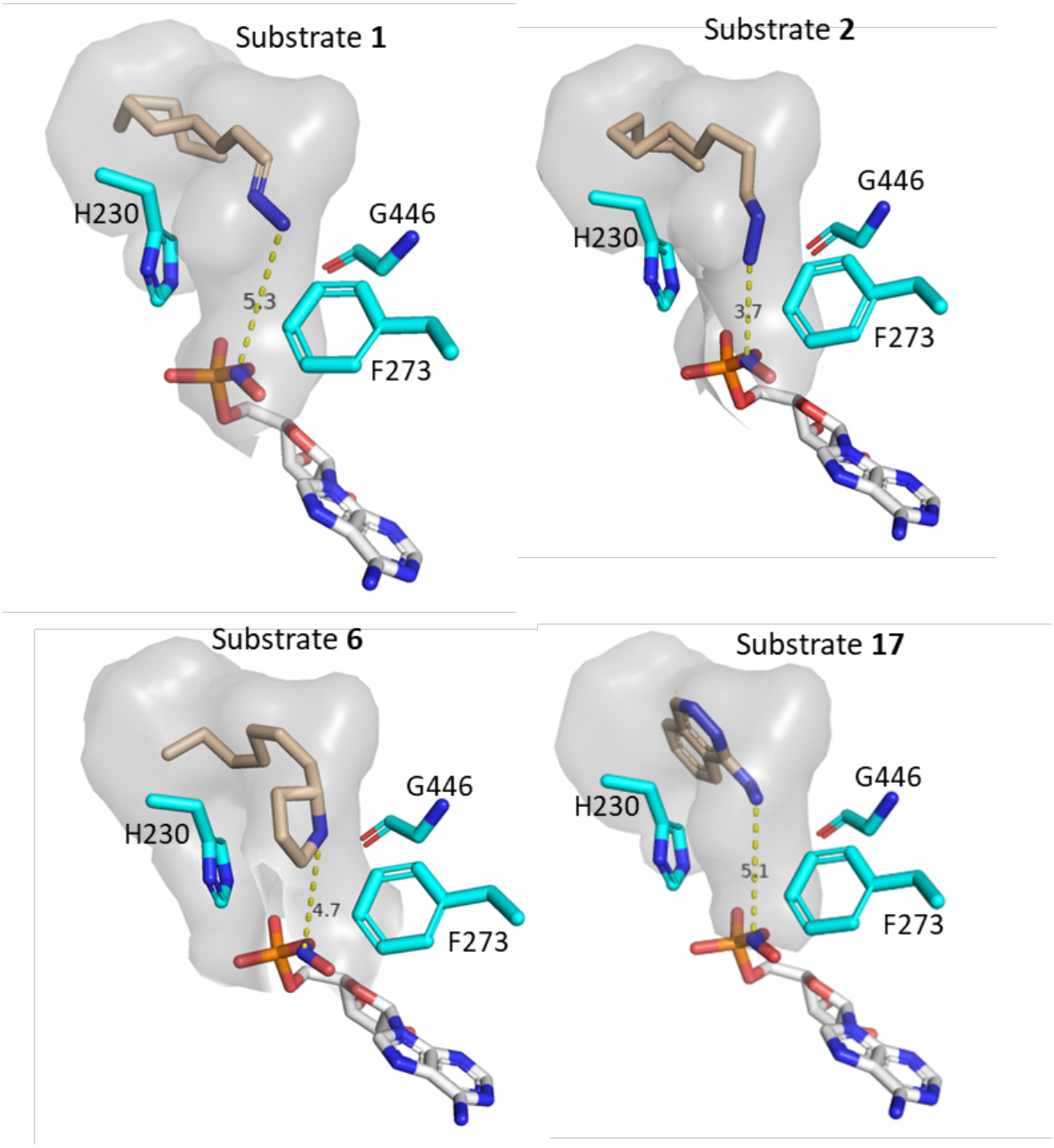
Fig. 6 Overall structure of Tri17 and mutagenesis results of Tri17. a) Crystal structure representing the adenylation conformation of Tri17. Conserved residues in the ATP binding pocket. ATP and Mg^{2+} were manually built into the apo structure via alignment with Nt4CL2[41] (PDB: 5BSM). b) Graph representing the relative activities of Tri17 variants compared to the wild-type enzyme when utilizing **1** as a substrate. The data points and error bars represent the average and standard deviations from three independently performed experiments, respectively. c) Potential binding pocket for various substrates. The pocket is constructed by residues from the *N*-terminal (light blue) and *C*-terminal (pale green) domains, shaped with a narrowed neck forming a tunnel toward the nitroso-AMP.



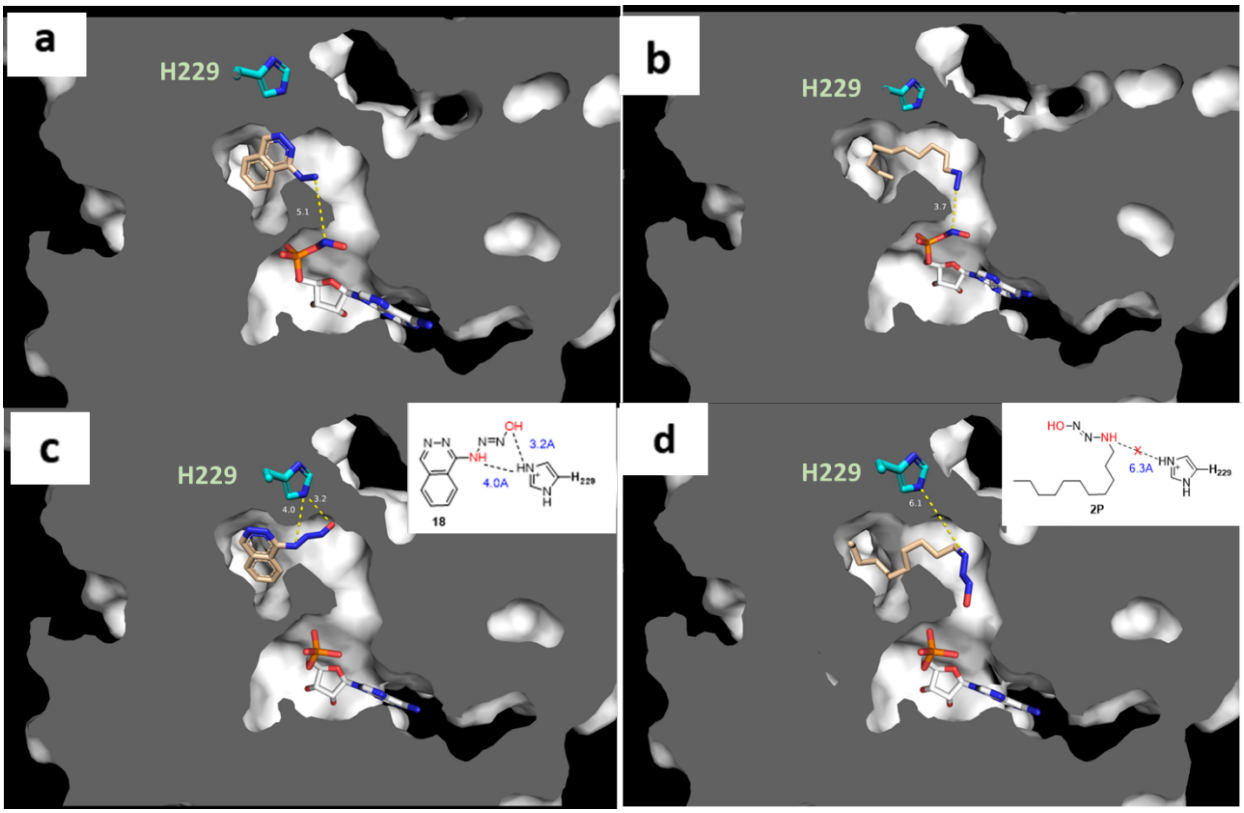
Extended Data Fig. 1 Biochemical analysis of Tri17 assays with 20. EICs demonstrating production of **21** and **22** from assays containing Tri17, ATP, nitrite, and **20**. Omission of any of these components led to the abolition of **21** and **22**. Utilization of ^{15}N-nitrite resulted in the expected mass spectral shifts for both **21** and **22**. No new products were detected when Aha11 was used in place of Tri17. A 10-ppm error mass tolerance was used for each trace. At least three independent replicates were performed for each assay, and representative results are shown.



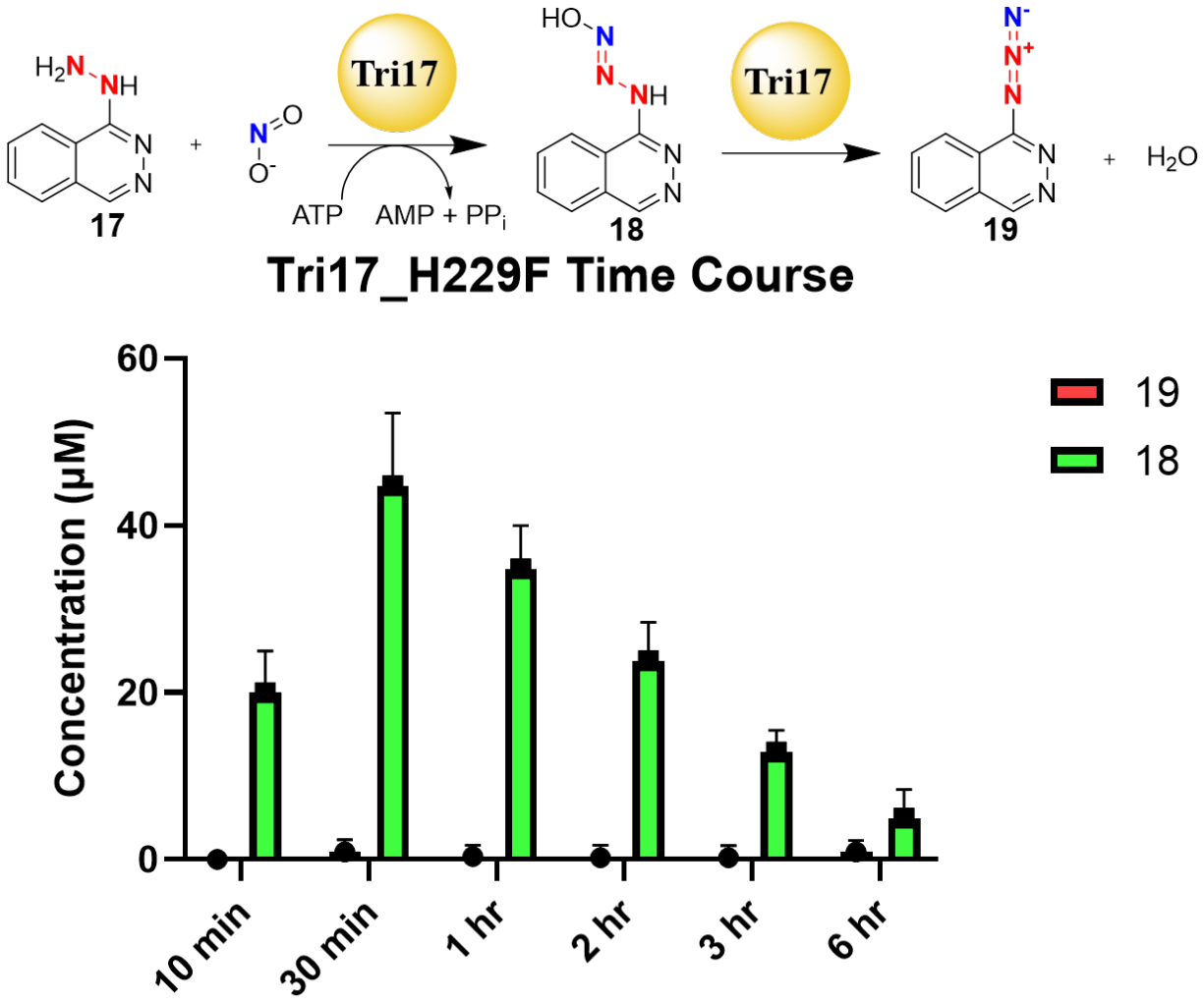
Extended Data Fig. 2 Biochemical analysis of Tri17-mediated dehydration of 18. A Tri17 biochemical assay with **17** was incubated at room temperature for 30 minutes and the protein was removed immediately using an Amicon spin filter (2 kDa MWCO). The reaction flowthrough was extracted with ethyl acetate, dried, and served as substrates (containing a mixture of **17**, **18**, and **19**) for new Tri17 reactions and the production of **19** was monitored after 3 hr. The EICs demonstrate increased production of **19** after a three-hour incubation period in a Tri17-dependent manner. A 10-ppm error mass tolerance was used for each trace. The data points and error bars present in the bar graph represent the average and standard deviations of **19** produced from three independently performed experiments.



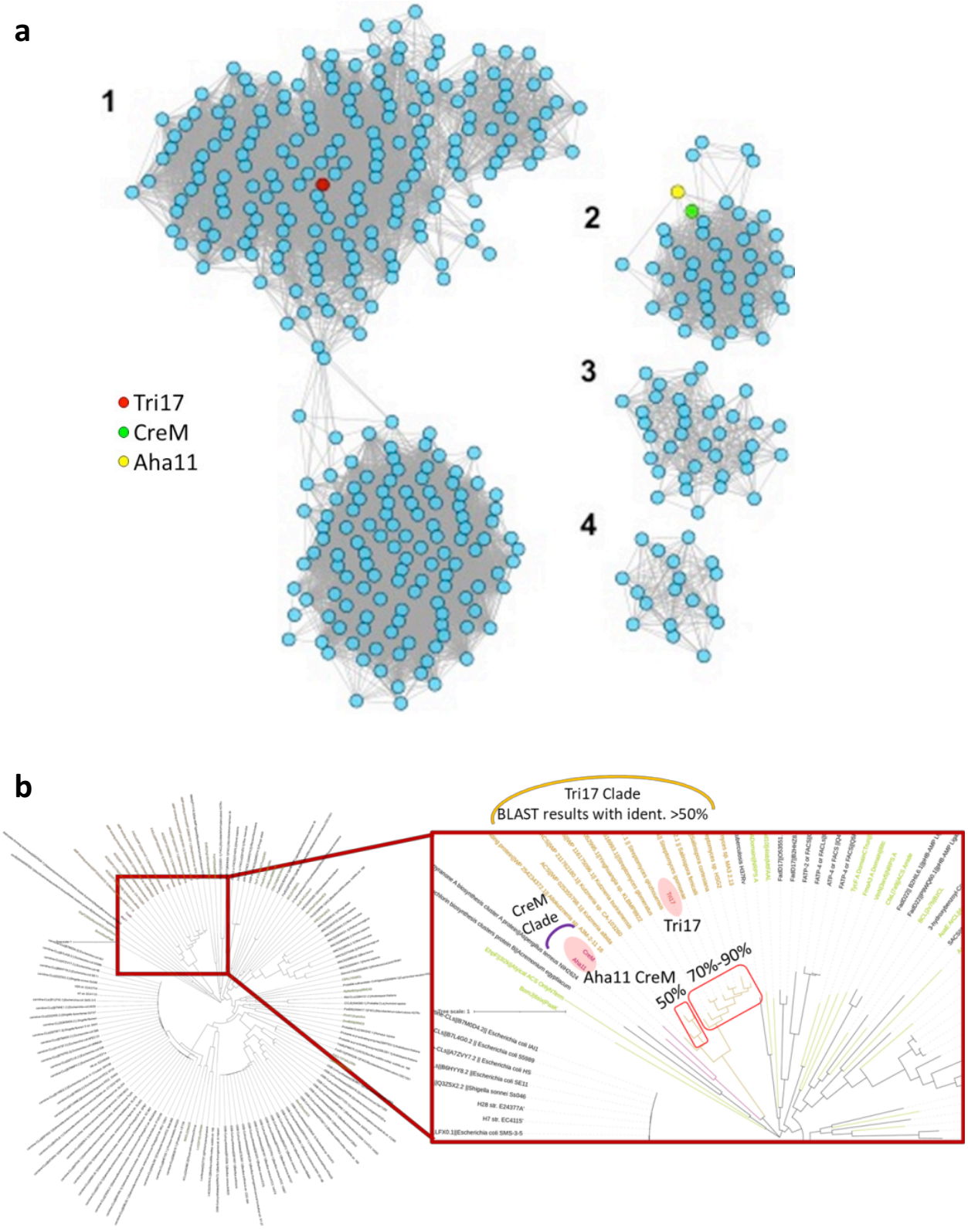
Extended Data Fig. 3 Potential binding pocket of Tri17 in Con_{Nuc}. Substrates **1**, **2**, **6** and **17** are docked into the potential pocket, exhibiting a binding mode in which the nucleophilic functional group will reach nitroso-AMP through the narrow neck formed by G446, F273 and H230 (colored cyan).



Extended Data Fig. 4 Comparison of the binding mode reveals a potential catalytic residue H229 functioning as a catalytic acid/base that participates in dehydration of 18. The azido-forming substrate **17,** linear substrate **2**, and their potential intermediate **18** and **2P** are docked into the binding pocket and shown as **a**), **b**), **c**) and **d**), separately. The bulkier moiety of **17** would be accommodated in the larger part of the pocket, leading to the N-N-N-O moiety approaching H229 (as shown in **a** and **c**). The linear substrate would extend more toward AMP, making the N-N-N-O moiety away from H229 (shown in **b** and **d**).

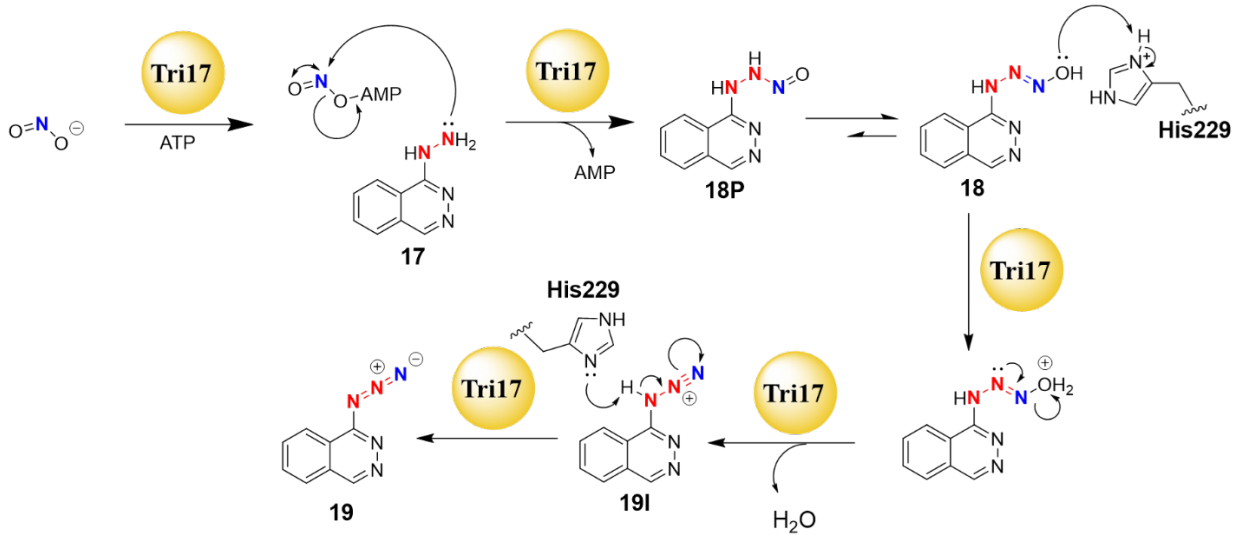


Extended Data Fig. 5 Biochemical analysis of Tri17_H229F with 17. Relative amounts of **18** and **19** quantified by LC-HRMS over a 6-hour time course from assays containing Tri17_H229F, ATP, nitrite and **17**. Error bars correspond to standard deviation of the mean from three replicate experiments.



Extended Data Fig. 6 Bioinformatic analysis of Tri17 and its homologs. a) Sequence similarity network (constructed using the EFI-Enzyme Similarity Tool using default

settings[85,86]) consisting of 1,471 Tri17 homologs represented as nodes. Each node represents proteins that are >45% identical. Tri17 is highlighted in red as part of Group 1 (1,159 members). CreM and Aha11 are highlighted in green and yellow, respectively, as part of Group 2 (191 members). Groups 3 and 4 are composed of 76 and 45 members, respectively. b) Phylogenetic tree of Tri17 suggests that Tri17 is located at a different clade than CreM and Aha11 (see **Supplementary Fig. 30** for larger representation of phylogenetic tree). The Tri17 clade shown in yellow was putatively annotated to include homologs with >50% sequence identity. The CreM clade is shown in purple consisting of CreM and Aha11. Structural homologs of Tri17 from the Dali server are colored green. Other proteins in black correspond to BLAST results with less than 50% sequence identity with respect to Tri17. AvaA6, the known Tri17 homolog, was not included in this analysis as its sequence was not available from the relevant publication[26].



Extended Data Fig. 7 Proposed reaction mechanism for azidation of 17 by Tri17. Tri17 utilizes ATP to activate nitrite to generate a nitroso-AMP intermediate that is subjected to nucleophilic attack by **17**. After tautomerization, compound **18** undergoes dehydration presumably through acid-base catalysis mediated by His229 to generate **19**.