Molecular basis of enzymatic nitrogen-nitrogen formation by a family of zinc-binding cupin enzymes

Guiyun Zhao1,2,5, Wei Peng3,5, Kaihui Song2, Jingkun Shi2, Xingyu Lu4, Binju Wang3✉ & Yi-Ling Du1,2✉

Molecules with a nitrogen-nitrogen (N-N) bond in their structures exhibit various biological activities and other unique properties. A few microbial proteins are recently emerging as dedicated N-N bond forming enzymes in natural product biosynthesis. However, the details of these biochemical processes remain largely unknown. Here, through in vitro biochemical characterization and computational studies, we report the molecular basis of hydrazine bond formation by a family of di-domain enzymes. These enzymes are widespread in bacteria and sometimes naturally exist as two standalone enzymes. We reveal that the methionyl-tRNA synthase-like domain/protein catalyzes ATP-dependent condensation of two amino acids substrates to form a highly unstable ester intermediate, which is subsequently captured by the zinc-binding cupin domain/protein and undergoes redox-neutral intramolecular rearrangement to give the N-N bond containing product. These results provide important mechanistic insights into enzymatic N-N bond formation and should facilitate future development of novel N-N forming biocatalyst.

1 State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University, 310003 Hangzhou, China. 2 Institute of Pharmaceutical Biotechnology, School of Medicine, Zhejiang University, 310058 Hangzhou, China. 3 State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, 361005 Xiamen, China. 4 Key Laboratory of Precise Synthesis of Functional Molecules of Zhejiang Province, School of Science, Instrumentation and Service Center for Molecular Sciences, Westlake University, 310024 Hangzhou, China. 5 These authors contributed equally: Guiyun Zhao, Wei Peng. ✉email: wangbinju2018@xmu.edu.cn; yldu@zju.edu.cn

https://doi.org/10.1038/s41467-021-27523-x
Molecules containing nitrogen-nitrogen (N-N) linkages display diverse structures and biological activities and make up a large proportion of clinical drugs. Although hundreds of these molecules have also been isolated from nature (Fig. 1a), the biosynthetic enzymes that are responsible for constructing these unusual N-N bonds, including hydrazine, diazo, and N-nitroso moieties, are only starting to be uncovered very recently. Among them, the piperazate synthase KtzT is a heme-dependent enzyme that catalyzes N-N cyclization of \(L-N^5\)-OH-ornithine to give the hydrazine-bearing piperazate, which is a nonproteinogenic amino acid building block for many non-ribosomal peptides. In the biosynthetic route to the cancer chemotherapeutic streptozocin, an iron-binding metalloenzyme SznF appears to mediate an oxidative rearrangement of \(N\)-methyl-\(L\)-arginine to give an \(N\)-nitrosoureia product. Among several enzymes from distinct protein families that have been related to N-N bond formation, the catalytic details of these unusual biochemical transformations are largely unknown in most cases.

It has been noticed that a collection of three genes, encoding a lysine/ornithine \(N\)-hydroxylase, a cupin protein, and a methionyl-tRNA synthase (MetRS) homolog, are widely distributed in bacterial species. In some cases, the cupin and MetRS-like proteins are fused into a single di-domain protein. However, it was until recently, the functions of these genes have been linked to the biosynthesis of N-N bond containing microbial specialized metabolites through in vivo studies. For instance, the di-domain proteins, consisting of an N-terminal cupin domain and a C-terminal MetRS-like domain, have been identified in the biosynthetic gene clusters (BGCs) of molecules including s56-p1, triacsin A, and pyrazomycin (also known as pyrazofurin). Preliminary studies based on in vivo bio-transformation experiments have shown that Spb40 from the s56-p1 BGC catalyzes hydrazine bond formation between \(L-N^6\)-OH-lysine and \(L\)-glycine, whereas PyrN from the pyrazomycin pathway displays alternative substrate specificity, linking \(L-N^6\)-OH-lysine to \(L\)-glutamate to form the product. However, due to the lack of in vitro biochemical data, how this family of hydrazine synthases mediates N-N bond formation remains elusive.

In this study, through in vitro biochemical assays, catalytic intermediate characterization, enzyme mutagenesis studies, and computational simulations, we reveal the detailed reaction route and the catalytic mechanism of this unusual N-N bond formation process, which involves a family of zinc-binding cupin proteins/domains. Our results provide important mechanistic insights into the biosynthetic strategies for N-N bond construction in nature.

**Results**

**In vitro reconstitution of PyrN-catalyzed N-N bond formation.** To reveal the details of PyrN-mediated N-N bond formation, we set out to reconstitute the PyrN-catalyzed reaction in vitro. We first prepared the N-terminal His-tagged PyrN protein from the E. coli heterologous expression system (Supplementary Fig. 1a). Considering that PyrN is predicted to contain a C-terminal MetRS-like domain that shares sequence homology to aminoacyl-tRNA synthetases (AARSs) (Fig. 2a), we incubated the isolated PyrN protein with chemically synthesized \(L-N^6\)-OH-lysine (2) and \(L\)-Glu, in the presence of ATP, Mg\(^{2+}\), and glutamyl-tRNA, the latter of which was provided in S30 premix extract (Promega).

---

**Fig. 1** Selected natural products containing a nitrogen-nitrogen bond and enzymatic N-N bond formation by a family of di-domain proteins. **a** Selected N-N bond containing microbial natural products. **b** Reaction catalyzed by the di-domain enzyme Spb40/Tri28 or PyrN in the biosynthetic pathway of s56-p1, triacsin A, and pyrazomycin, respectively.
Moreover, total tRNA mixtures from E. coli (Roche) was also used in place of glutamyl-tRNA in a separated assay. LC-MS analysis of the above reaction mixtures after Fmoc chloride (Fmoc-Cl) derivatization, revealed that they both generated a new product with the same retention time and mass signal (m/z 514, [M + H]^+) ion to that of Fmoc-1 (Fig. 2b), the latter of which was isolated from the culture supernatant of an engineered E. coli strain expressing both the lysine-N^δ-hydroxylase gene nbtG and pyrN^δ, and structurally characterized based on extensive NMR spectroscopic analysis (Supplementary Fig. 2). The N-N linkage in Fmoc-1 was confirmed by using 15N NMR and 1H,15N heteronuclear multiple bond correlation (HMBC) NMR analysis, on the Fmoc-1 sample prepared with 1-ε-15N-lysine and 1,15N-glutamate as feeding precursors (Supplementary Fig. 2g, h). Together, these results demonstrated that 1 was successfully produced in the above in vitro reaction mixtures.

Next, we test the tRNA-dependence of PyrN-catalyzed N-N bond formation. We removed glutamyl-tRNA/total tRNA mixtures from the above in vitro reaction mixtures or added RNase in the assay. Interestingly, 1 was still produced in all these cases, suggesting that the PyrN-mediated reaction is tRNA-independent (Fig. 2b and Supplementary Fig. 3a). To further interrogate the necessity of other components in the above in vitro system, they were individually eliminated from the reaction mixtures, and followed by LC-MS analysis. The results demonstrated that 1-Glu, 2, PyrN, ATP, Mg^{2+} are essential to the production of 1 (Fig. 2c).

The catalytic role of MetRS-like domain in PyrN-catalyzed reaction. We next set out to identify potential reaction intermediate(s) in PyrN-catalyzed N-N bond formation. Based on the sequence homology of PyrN MetRS-like domain to aminoaacyl-tRNA synthetases (AARSs), as well as the ATP-dependence of PyrN-catalyzed reaction, we speculated that the MetRS-like domain could catalyze the initial ligation of 1-Glu and ATP to form the proposed 1-aminoadenyl-1-Glu intermediate. The essential roles of both MetRS-like domain and cupin domain in PyrN-catalyzed reaction. The di-domain organization of PyrN and the previous in vivo studies on Spb40/PyrN, have suggested that the formation of N-N linkage is likely a multi-step process^3,8. Having established the in vitro reaction condition for the full-length PyrN, we next explore the catalytic role of each domain in vitro. The ATP-dependent nature of PyrN-catalyzed reaction is consistent with the previous in vivo results, which shows that a point mutation at Lys458 in the putative ATP-binding motif of MetRS-like domain abolished the activity of PyrN^δ. In line with this result, PyrN (K458A) also showed negligible activity in vitro (Fig. 2d). We next interrogate the role of the N-terminal cupin domain (Fig. 2a). The cupin family proteins are known for their ability to bind metal ions for mediating a variety of biochemical transformations^9. Here, we mutated the four conserved putative metal-chelating residues (His50, His52, Glu56, and His90) in the cupin domain (Fig. 2a), and assayed the activities of the corresponding variants in vitro (Supplementary Fig. 3b). We found that the E56A variant completely abolished the production of 1, whereas the H50A and H90A variants still retain significant activities (~82 and ~27% relative to the wild-type enzyme, respectively) (Supplementary Fig. 3b and 3c). In addition, the activities of the H52A and the H50A/H90A double variants were severely affected (<3%). Taken together, these results demonstrated that both the MetRS-like domain and the cupin domain are indispensable for PyrN-catalyzed reaction.

Next, we determine whether the two domains of PyrN act independently. We observed that, although both the E56A and K458A variants are inactive on their own, the production of 1 can be restored in a reaction mixture with both variants presented (Fig. 2d). Furthermore, 1 can also be generated upon incubation of the truncated PyrN protein (Δ1–108 aa) that only carries the MetRS-like domain, with the one (Δ110–668 aa) that only contains the cupin domain (Fig. 2a, e). Together, the above results established the essential roles of both MetRS-like domain and cupin domain in PyrN-mediated transformations to afford 1.
afford glutamyl-AMP, which is in analogy to the first half reaction of AARS20. The subsequent combination of glutamyl-AMP with 2 could either be mediated by both the MetRS-like and cupin domains, or the cupin alone. To unveil the reaction details, we first focus on the PyrN (E56A) variant, which contains a functional MetRS-like domain, and has a higher protein expression level compared to that of the truncated MetRS-like domain alone (Supplementary Fig. 1a). Incubation of the PyrN (E56A) variant enzyme with L-Glu, 2, ATP, and Mg²⁺ afford a product (3) that displays the same mass signal (m/z 292) but different retention time to that of 1 (Fig. 3a). Moreover, the production of 3 is also strictly dependent on all the components that are essential for the wild-type PyrN reaction (Fig. 3a), and both these reactions generate AMP (Supplementary Fig. 4). We further found that 3 can also be produced when we substituted the truncated MetRS-like domain protein PyrN (∆1-108 aa) for the PyrN (E56A) variant protein in the above assay (Fig. 3a). Together, these results suggested that 3 might be the reaction product from the PyrN MetRS-like domain.

To characterize compound 3, we first prepared 18O-labeled or 13C₆-labeled substrate 1-N6,18O-H-Lys (18O-2) and 1-N6,13C₆-Lys (13C₆-2), and included them in the in vitro assays of both wild-type PyrN and its E56A variant. Subsequent LC-HR (high resolution)-MS/MS analysis of the above reaction mixtures demonstrated that, although 3 and 1 share the same molecular weight, their fragment patterns are distinct from each other (Fig. 3b, c and Supplementary Fig. 5). More specifically, the 18O atom is still attaching to the N6 atom of the lysine structural subunit in 3, whereas the 13C₆ atom in 1 has been cleaved from the N6 atom and rearranged into one of the carboxyl oxygens in the glutamate subunit (Fig. 3c). This fragment pattern of 3 is supportive of an amide molecule, N-glutamyl-N6-hydroxyl-lysine. To further ascertain the structure of 3, we chemically synthesized the authentic compound of N-glutamyl-N6-hydroxyl-lysine, which displays identical retention time and fragment pattern to that of 3 in LC-MS/MS analysis (Fig. 3a and Supplementary Fig. 6). Together, the above results unambiguously established the structure of 3 as N-glutamyl-N6-hydroxyl-lysine.

We next test whether 3 is an on-pathway reaction intermediate. We incubated 3 with PyrN, however, 1 was not detected in the reaction mixture upon LC-MS analysis (Supplementary Fig. 7), indicating that 3 is likely an off-pathway shunt product that might derive from an unstable reaction intermediate. To get access to the direct enzymatic product from the MetRS-like domain, we quickly quench the E56A reaction shortly after the reaction is initiated, through the removal of proteins by ultracentrifugation, and immediately analyze the filtrate by LC-MS. We found that, besides 3, another small peak (4) with the same mass signal (m/z 292) could be detected (Fig. 3d). Time course study revealed that this unstable compound rapidly converts to 3, with a half-life estimated to be less than 10 min.

It has been known that O-acyl hydroxylamine rearranges to N-acyl hydroxylamine19,21. For instance, in the biosynthetic pathway of antibiotic valanimycin, an unstable ester intermediate with a similar N-O bond nonenzymatically converts to the corresponding amide shunt product21. Accordingly, our results strongly suggested that the genuine product from the MetRS-like domain is the unstable ester O-glutamyl-N6-hydroxyl-lysine, which rapidly arranges to 3 in the absence of a functional cupin domain (Fig. 3e). This proposal was also supported by the above 18O-labeling pattern in product 1, which indicated the presence of such an ester intermediate in the PyrN-catalyzed N-N bond formation (Fig. 3c). A similar putative ester intermediate was also proposed in the Spb40-mediated reaction5. We further performed LC-HR-MS/MS analysis on 4, which was generated in situ by the E56A-catalyzed reaction (Supplementary Fig. 8). The results again supported the structural assignment of 4 as O-glutamyl-N6-hydroxyl-lysine. Altogether, our results demonstrated that the PyrN MetRS-like domain catalyzes ATP-dependent condensation of L-Glu and 2 to give 4, which might subsequently undergo cupin-mediated intramolecular arrangement to afford the N-N-containing product 1.

The catalytic role of cupin domain in PyrN-catalyzed reaction. We next investigate cupin-catalyzed N-N bond formation. Our previous study has suggested that the cupin domain of PyrN

Fig. 3 Characterization of the reaction intermediate and shunt product from the PyrN-catalyzed reaction. a LC-MS analysis of the in vitro reaction mixtures of the PyrN (E56A) variant and the truncated MetRS-like domain of PyrN, PyrN MetRS. b, c LC-HR-MS/MS analysis of 1 and 3 in the reaction mixtures containing isotope-labeled substrates. Selected diagnostic fragment signals are displayed in (c). d Detection of an unstable reaction intermediate 4 by LC-MS analysis. e Proposed reactions routes of PyrN-catalyzed N-N bond formation based on this study.
binds zinc ion\(^8\). Here, we use ICP-MS (inductively coupled plasma-mass spectrometry) to further analyze the artificially truncated cupin domain, PyrN (Δ110–668 aa), revealing ~0.5 equivalent of bound zinc. We then test the metal-dependence of PyrN-catalyzed N-N bond formation by the inclusion of metal-chelating agent 1,10-phanthrolnine (OP) or ethylenediaminetaetic acid (EDTA) into the in vitro assays. We found that OP could efficiently inhibit the production of 1 in a PyrN-catalyzed reaction (Supplementary Fig. 9a). More specifically, OP only blocks the N-N bond-forming step mediated by cupin domain, as we still could detect the production of 3 (Supplementary Fig. 9b).

On the other hand, although EDTA appeared to have no effect on the wild-type PyrN (Supplementary Fig. 9a), it completely inhibits the N-N forming activity of the H90A and H50A variants (Supplementary Fig. 9c), indicating that the metal ion is relatively tightly bound by PyrN, but could be stripped out by EDTA after one of metal-chelating His residues was replaced. Taken together, the above results established the metal-dependent nature of cupin-catalyzed N-N bond formation.

Identification of a zinc-binding N-N forming cupin enzyme by genome mining. Our attempts to crystalize either the full-length PyrN or its artificially truncated versions have been fruitless so far, we thus turn to search for naturally occurring standalone cupin proteins that might have the same catalytic activity as that of the PyrN cupin domain. We use a genome mining approach to specifically target putative BGCs that also carry homologs of the N\(^6\)-lysine hydroxylase gene pyrM and the putative saccharopine dehydrogenase gene pyrL, the latter of which was suggested to convert 1 to 2-hydrizinoglutaric acid in the biosynthetic pathway of pyrazomycin\(^17,18\). Based on this mining strategy, a putative BGC from strain Rhodococcus jostii RHA1 was identified and drew our particular attention (Supplementary Fig. 10a). The standalone cupin protein (NCBI accession number: WP_007299751, renamed here as RHS1) from this BGC has been previously crystallized and shown to be a zinc-binding protein (PDB accession number: 5UQP) (Fig. 4a and Supplementary Fig. 11). However, neither the metabolic product from this putative BGC nor the function of the cupin protein RHS1 were known. We generated an engineered Rhodococcus jostii strain that overexpressed its own N\(^6\)-lysine hydroxylase, the standalone MetRS-like protein, and RHS1. LC-MS analysis of the culture supernatant of this strain revealed the production of 1, supporting that the standalone MetRS-like protein also utilizes 2 and l-Glu as substrates (Supplementary Fig. 10b). To further confirm whether RHS1 is functionally equivalent to the PyrN cupin domain in vitro, we prepared His\(_6\)-tagged recombinant RHS1 (Supplementary Fig. 1b) and included it in the in vitro assays with the PyrN MetRS-like domain protein or the PyrN (E56A) variant, both of which could generate the unstable product 4 in situ for the subsequent cupin-mediated reaction. LC-MS analysis of the above reaction mixtures revealed the production of 1, demonstrating that RHS1 has the same N-N bond-forming activity to that of the PyrN cupin domain (Fig. 4b).

Analysis of the RHS1 crystal structure revealed that the zinc ion is coordinated by Asp63, His65, Glu69, and His103 (Fig. 4a). We individually replaced these residues with alanine and assayed the resulting variants. LC-MS analysis revealed that all these variants failed to convert 4 to 1 (Fig. 4c). We then determined the zinc contents of these variants by using ICP-MS, which showed that both the H65A and H103A variants lost their zinc-binding abilities, whereas the D63A and E69A variants still bind one equivalent of zinc ion (Supplementary Fig. 13a). We further found that all these four RHS1 variants failed to restore their activity even in the presence of exogenous zinc ion (100 μM) in the reaction buffer (Supplementary Fig. 13b). Beside the above zinc-binding residues, we also mutate residues surrounding the active site cavity. These includes Arg26, Ser28, Leu32, Ile52, Thr60, Val67, Trp71, and Trp119 (Fig. 4a). We observed that, except the three variant genes (I52A, T60A, and W71A) that were expressed in the inclusion body, all the other soluble enzymes could still produce significant amounts of 1 in in vitro assays, indicating that they are not critical for catalysis (Supplementary Fig. 13c).

Computational studies of RHS1-catalyzed N-N bond formation. The high instability of substrate 4 makes it impossible to obtain an enzyme-substrate co-crystal structure, we thus turn to computational approaches to gain mechanistic insight into the novel cupin-catalyzed N-N bond formation. On basis of the RHS1 crystal structure, 4 was docked into the active site (Supplementary Fig. 14). Two representative binding conformations (conformation-1 and conformation-2) were selected for further mechanistic study (Supplementary Fig. 14a, b). Molecular dynamics (MD) simulations showed that both conformations are stable (Supplementary Fig. 14c, d). For conformation-1, our QM calculations show that the first N-O cleavage step is endothermic by 41.4 kcal/mol (RC\(^{-}\)→ IC1\(^{-}\) in Supplementary Fig. 15), indicating that conformation-1 is highly unfavorable for N-O cleavage reaction and thus can be reasonably ruled out. We then focus on conformation-2, in which the substrate O5 and N1 atoms are coordinated to Zn (Supplementary Fig. 14b). The QM/MM calculated energy profile for the RHS1-catalyzed conversion of 4 to 1, from the MD equilibrated representative structure of conformation-2, is shown in Fig. 5a. In the initial reactant complex (RC), the imino group (N1-H) of 4 forms a strong H-bond with the carboxyl O2 of Glu69, while the NH3 group of 4 is H-bonded to the carboxyl O4 of Asp63 (Fig. 5b and Supplementary Fig. 16a).

**Fig. 4** Identification of RHS1 as a zinc-binding cupin enzyme catalyzes N-N bond formation. a The putative active site of RHS1 (reproduced from PDB: 5UQP) with the zinc ion presented as a gray sphere. b, c In vitro biochemical assays of RHS1 and its variants.

---

**Supplementary Information**

5UQP) with the zinc ion presented as a gray sphere.
Starting from RC, our QM/MM calculations show that the N1-O1 cleavage in 4 is coupled with the N1-O2 bond formation between the cleaved N1-containing fragment and the carboxyl group of Glu69 (Fig. 5a). This step experiences the barrier of 20.5 kcal/mol (16.9 kcal/mol in QM simulations in Supplementary Fig. 17a, b) and leads to the formation of N1-O1 cleaved intermediate IC1 (Fig. 5a and Supplementary Fig. 18). In order to initiate the following N-N coupling reaction, the NH3 group of 4 has to lose at least one proton and thus being deprotonated. Starting from IC1, we have investigated three possible reaction pathways for this deprotonation step, including the transfer of proton to: (I) carboxyl O4 of Asp63 (Supplementary Fig. 19), (II) carboxyl O3 of a substrate without bridging waters (Supplementary Fig. 20), and (III) carboxyl O3 of the substrate with the help of bridging waters (Fig. 5a). Among these pathways, we found that the two proton transfer pathways (I) and (II) are unfavorable thermodynamically, while the only feasible pathway is the proton transfer to O3 assisted by two bridging waters. As such, two water molecules and the terminal carboxyl group of the substrate were further included in the QM region (IC1') (Supplementary Fig. 16b). For such water-assisted deprotonation of NH3 via TS2, the reaction involves a small barrier of 5.8 kcal/mol (IC1' → TS2 in Fig. 5a), leading to the intermediate IC2. Our QM model calculations further confirm the vital roles of water molecules in facilitating proton transfer from amino to the carboxyl group (Supplementary Fig. 17a, c).

Staring from IC2, we found the direct N-N coupling is highly unfavorable kinetically. This is mostly because the NH2 group is relatively far away from N1-H and such conformation is highly unfavorable for N-N coupling. As such, MD simulations was carried out to identify if IC2 can experience some conformational change to facilitate the following N-N coupling (IC2 → IC2' in Fig. 5a). Interestingly, we found that the protonated carboxyl group (O3 site in Fig. 5b) can approach the deprotonated carboxyl group (O6 site in Fig. 5b) and thus forms a persistent H-bond between each other during the MD simulations (Supplementary Fig. 21). We also docked product 1 into RHS1 and found that 1 took a curly binding conformation that is quite similar to IC2' (Supplementary Fig. 22). All these findings further support the rationality of such conformational change (IC2 → IC2').
IC2’ in Fig. 5a). Moreover, we found that this conformational change could trigger the rotation of the N2-H group toward the non-enzymatic N1-O2 cleavage, en route to the N-N coupling product. However, the reaction requires a high barrier of 32.4 kcal/mol and is thus highly unfavorable (Supplementary Fig. 24). We also investigated the underlying mechanism for the nonenzymatic conversion of 4 to 3 using the cluster-continuum model calculations. Our calculations showed that the overall reaction likely involves three key steps, including the hydration of ester 4, C-O bond cleavage, and amide bond formation (Supplementary Fig. 25). These calculated results are consistent with our above experimental data, which shows that the conversion of 4 to 3 can proceed nonenzymatically, whereas the production of 1 from 4 only occur in the presence of cupin enzymes.

Our above calculations showed that Glu69 plays a key and unusual role in mediating N-N formation, which is in line with our results from the point-mutation experiments of RHS1 and PyrN, revealing that both the RHS1 (E69A) and PyrN (E65A) variants completely abolished their activities (Fig. 4c and Supplementary Fig. 3). We generated four more RHS1 variants (E69D, E69N, E69Q, and E69L) and then evaluated their protein expression level and in vitro catalytic activity. We found that the E69Q and E69N variants are also inactive, and both the E69D and E69L variants expressed exclusively in the inclusion body, preventing their further activity assays (Supplementary Fig. 26).

**Discovery of new PyrN homologs with alternative amino acid substrate specificity.** To facilitate the application of this hydrazine-forming enzyme family in biocatalysis, we search for new PyrN homologs that could potentially recognize other amino acid substrates, besides 1-glutamate and 1-glycine. We used Sequence Similarity Network (SSN) tool to analyze the phylogenetic relationships of PyrN homologs in the Uniprot protein database (Supplementary Fig. 27a), which guided our selection of more than ten genes (for the di-domain proteins) or gene pairs (for the standalone MetRS-like proteins and their associated cupin proteins) for gene synthesis and the subsequent catalytic activity evaluation (Supplementary Fig. 27b, c). Besides the enzymes that share the same substrate specificity with Sbe40/Tri28, which recognize 1-glycine, new homologs that can utilize 1-serine, 1-alanine, and 1-tyrosine were identified (Supplementary Figs. 28–31). Interestingly, we further found that some of cupin domains/proteins have relaxed substrate specificities, as they could be coupled with a unnative partner, the MetRS-like domain of PyrN, to produce 1 (Supplementary Fig. 32). Considering that 1 was not detected when these cupin domains/proteins were co-expressed with their native MetRS-like domains/proteins, these results suggested that MetRS-like domains/proteins have more strict substrate specificities, and could act as a gatekeeper to determine the outcome of the multi-step reactions. It is also worth to mention that all the cupins we investigated contain an invariant Glu residue that has been suggested to play a critical role in mediating N-N bond formation, based on our above computational studies (Supplementary Fig. 33).

**Discussion**

The biosynthetic strategies for constructing N-N bonds in natural products have received great attention over the past decade. Despite that there is an increasing number of dedicated enzymes from different protein families that were discovered to be responsible for the formation of N-N linkages in various natural products, little is known about their catalytic mechanism. In this study, we conducted detailed in vitro characterization of PyrN, a di-domain enzyme that catalyzes hydrazine bond formation in the biosynthetic pathway of antiviral antibiotic pyrazomycin. Our results revealed that the MetRS-like domain of this enzyme catalyzes an ATP-dependent condensation of 1-N6-OH-lysine and 1-glutamate to give a highly unstable ester intermediate, which could proceed through a glutamyl-adenylate intermediate, as observed in the first half reaction of canonical aminoacyl-tRNA synthetases (AARSs). However, unlike AARSs, MetRS-like enzymes appear to utilize 1-N6-OH-lysine, instead of tRNA, as an amino acid carrier in the second half reaction. In the presence of the zinc-binding cupin domain, the unstable ester product from the MetRS-like domain-catalyzed reaction undergoes redox-neutral, intramolecular arrangement to form an N-N bond containing product. This could be achieved through the zinc-assisted N-O bond cleavage followed by N-N bond formation, as suggested by our computational simulations. Our result thus demonstrated that these zinc-binding cupin domains/proteins represent a novel family of N-N-forming enzymes. Considering that this cupin-family enzymes are widespread in bacteria and hydrazine synthases have great potential in biocatalysis, we further performed SSN-guided protein database mining. Such effects have resulted in the successful identification of new homologs of either the full-length PyrN or its MetRS-like and cupin domains, which are able to accept alternative amino acid substrates, expanding the potential application of these interesting biocatalysts.

In conclusion, we reported the catalytic route and the reaction mechanism of N-N bond formation by a family of zinc-binding cupin enzymes, which are widely distributed in the biosynthetic gene clusters of various bacterial specialized metabolites. Our results expanded the current knowledge of enzymatic N-N bond formation and set the stage for the development of novel biocatalysts for the synthesis of useful molecules containing a hydrazine bond.

**Methods**

**General methods.** DNA primers (primer sequences are listed in Supplementary Table 1) were purchased from Tsingke Biological Technology. Reagents were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Cambridge Isotope Laboratories, New England BioLabs, Bio Basic Inc. DNA manipulations in *Escherichia coli* strains were carried out according to standard procedures. Ampicillin (100 μg ml⁻¹), apramycin (50 μg ml⁻¹), kanamycin (50 μg ml⁻¹), and spectinomycin (50 μg ml⁻¹) were used for the selection of recombinant *E. coli* strains.

**DNA primers (primer sequences are listed in Supplementary Table 1) were purchased from Tsingke Biological Technology. Reagents were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Cambridge Isotope Laboratories, New England BioLabs, Bio Basic Inc. DNA manipulations in *Escherichia coli* strains were carried out according to standard procedures. Ampicillin (100 μg ml⁻¹), apramycin (50 μg ml⁻¹), kanamycin (50 μg ml⁻¹), and spectinomycin (50 μg ml⁻¹) were used for the selection of recombinant *E. coli* strains.**
Isolation and characterization of Fmoc-PyrN. The E. coli strain harboring vector pCDFSuet-nbG (an N-lysine hydroxylase gene) and pET28a-pyrN was cultivated in M9 medium for protein expression and metabolites production. When the OD600 reached ~0.6, the cultures were supplemented with IPTG at a final concentration of 0.1 mM along with 0.1% (w/v) of ethanol for deprotection. The solution was concentrated under vacuum, and the resulting residue was dissolved in methanol, dried, and then re-dissolved in water. Subsequent Fmoc chloride (Fmoc-Cl) derivatization was performed as follows: the aqueous solution (15 mM) containing the protein was mixed with 50 mM of concentrated HCl for 2 h at room temperature, and then the solution was concentrated under vacuum, and the resulting residue was dissolved in methanol, dried, and then re-dissolved in water. Subsequent Fmoc chloride (Fmoc-Cl) derivatization was performed as follows: the aqueous solution (15 mM) containing the protein was mixed with 50 mM of concentrated HCl for 2 h at room temperature, and then the solution was concentrated under vacuum, and the resulting residue was dissolved in methanol, dried, and then re-dissolved in water. The supernatant was collected and centrifuged at 8,000 ×g for 10 min. The supernatant was then subjected to LC-MS analysis with or without Fmoc-Cl derivatization. LC-MS analysis was carried out on an Agilent 1260 II-6125 apparatus, using an Agilent Eclipse XDB-C18 column (5 μm, 4.6 mm ID × 250 mm). Two elution methods were used for the separation of the two samples (buffer A: 0.1% formic acid + 0.1% acetic acid in water, buffer B: 0.1% formic acid + 0.1% acetic acid in acetonitrile). The gradient elution was performed at 0.5 mL min−1 with a mobile-phase gradient consisting of a linear gradient of water and acetonitrile (v/v): 85:15, 0–20 min; 59:41, 20–25 min, both of which contain 0.05% (v/v) formic acid (detection wavelength: 263 nm). For the detection of molecules without derivatization, elution was performed at 0.5 mL min−1 with a mobile-phase gradient consisting of a linear gradient of water and acetonitrile, both of which contain 0.05% (v/v) formic acid (detection wavelength: 210 and 254 nm). LC-HR-ESI-MS was performed similarly on waters UPLC (Waters Corp., USA) coupled with an AB TripleTOF 6500plus mass spectrometer system (AB SCIEX, USA).

Enzymatic synthesis of l-N6-18O-Lys and l-N6-18O-13C2-Lys. For the preparation of l-N6-18O-Lys (13C2O-2), a solution (3 mL) consisting of 2 mM l-Lys, 2 mM NADH in 40 mM Tris-HCl buffer (pH 8.0), and a solution (3 mL) consisting of 10 mM PyrM (N-lysine hydroxylase) in 40 mM Tris-HCl buffer (pH 8.0), were thoroughly bubbled with Argon separately to remove molecular oxygen. Subsequently, the 13C2O-2 was introduced to these sealed bottles and followed by mixing these solutions together to initiate the PyrM-catalyzed reaction. This reaction mixture was monitored by LC-MS analysis to evaluate the conversion of l-Lys to 13C2O-2, and the enzyme was removed through ultracentrifugation using Amicon Ultra-4 mL centrifugal filters (Millipore, 3,000 MWCO), when l-Lys completely converts to 13C2O-2. The filtrate solution containing 13C2O-2 was directly used for subsequent experiments without further purification.

Setup for computational studies. The initial structure of the enzyme was prepared on the basis of the determined crystal structure of the cupin protein from Rhodococcus jostii RHA1 (PDB code: 5UQP, determined at a resolution of 2.4 Å), which was retrieved from the Brookhaven Protein Data Bank (http://www.rcsb.org/pdb). Here, we assigned the protonation states of titratable residues (His, Glu, Asp) based on pKa values from the PROPKA software36 in combination with a careful visual inspection of the local hydrogen-bonding networks. Then all the residues in the substrate binding pocket except for the histidine (His) residues in the substrate binding pocket were replaced by alanine. The enzyme was then solvated in a rectangular box with TIP3P37 water, with a minimum distance of 15 Å from the protein surface. After the setup of the system, it was minimized by the steepest descent and conjugate gradient methods. The system was then energy minimized with the Amber 18 package38,39. All the system coordinates were restrained using the SHAKE, and the integration step of 2 fs was used. All the MD process were conducted by the Amber 18 package34.

Quantum calculation. The quantum model is comprised of the substrate, zinc cofactor, as well as zinc-coordinated residues, including Asp63, His65, Glu69, and His103. All quantum model calculations were performed with the Gaussian 16 software40. The geometry optimizations and frequency calculations were performed in the continuum solvation model at the B3LYP/def2-SVP level of theory. B3LYP was proven to be a successful functional for studying zinc-based metalloenzyme37,39.
The energies were further refined with the larger basis set de2-TZVP for all atoms. As for the nonenzymatic reaction, the species with hydrated cluster model22,23 were optimized in conjunction with the SMD continuum solvent model at the BMK/6-31 G(d) level of theory. The energies were further refined with the larger basis set 6-311 + G(d,p) for all atoms. The dispersion energies were included in both optimizations and single-point energy calculations. The OriginPro Learning Edition (https://www.originlab.com/OriginProLearning.aspx) was used to process the data from MD simulation, QM/MM and QM scan.

**QM/MM calculation.** For the subsequent QM/MM calculations, we selected the representative snapshot from each classical MD trajectory. All the QM/MM calculations were performed by the ChemShell,10 in which the turbomole<sup>11</sup> is invoked for the QM region while the DL_POLY<sup>12</sup> is used for the MM region. Besides, the electronic embedding scheme<sup>13</sup> was employed to account for the polarizing effect of the enzyme environment on the QM region, while the hydrogen link atoms with the charge-shift model was used to deal with the QM/MM boundary. Here, the QM region was studied with the hybrid B3LYP<sup>14</sup>,15 density functional, which has been proven to be reliable for the simulation of zinc-containing metalloenzymes<sup>16</sup>,17,18,19,20. In this study, the double-ζ basis set de2-SVP were used for geometry optimization, while the energies were corrected with the larger basis set de2-TZVP for all the QM region atoms. The dispersion corrections with Grimme’s D3 method<sup>21</sup> were added in all QM calculations. Similar to QM calculations, the QM region in the QM/MM model included the Zn atom, the substrate, the coordinated Asp63, His65, Glu69, and His103 (See Supplementary Fig. 16 for the selection of QM region). For the transition states (TSs) optimizations, we firstly performed the relaxed potential energy surface (PES) scanning. Then, the located highest point of PES was used to find the full TS optimizations using the DL-FIND code.<sup>22</sup>

**Characterization of PyrN homologs selected from Uniprot protein database mining.** The genes or gene pairs, obtained through SSUI-guided database mining (Supplementary Fig. 27), were subjected to codon optimization and gene synthesis (at Sangon Biotech Co., Ltd.). For fused di-domain enzyme genes, they were cloned into pET28a vector through the NdeI/Xhol sites, and co-introduced into the E. coli strain BL21(DE3) with pcFDueto-nbgR (an N<sup>ε</sup>-lysine hydroxylase gene) for protein expression and metabolite production. For standalone cupin and MetRS-like proteins, the MetRS-like genes were cloned into pET28a through the NdeI/Xhol sites, and the cupins were cloned into the NdeI/Xhol sites of pcFDueto-nbgR to afford pcFDueto-nbgR-cupin vectors. The two vectors (pET28a-metRS and pcFDueto-nbgR-cupin) were then co-introduced into the E. coli for protein expression and metabolism production. For IC-MS analysis, the culture broth supernatants of each strain were first mixed with two volumes of acetonitrile and followed by Fmoc chloride derivatization. The MS signals corresponding to the twenty Fmoc-lysine-AAs (amino acids) conjugates were searched for all the samples, to determine the substrate specificities of synthesized enzymes.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability** The data generated in this study are provided in this Article, Supplementary Information, and Source data files. A reporting summary for this article is available as a Supplementary Information file. All other relevant data are available from the corresponding author upon request. Source data are provided with this paper.

Received: 2 June 2021; Accepted: 19 November 2021; Published online: 10 December 2021

**References**

1. Blair, L. M. & Sperry, J. Natural products containing a nitrogen–nitrogen bond. *J. Nat. Prod.* 76, 794–812 (2013).
2. Du, Y.-L., He, H.-Y., Higgins, M. A. & Ryan, K. S. A heme-dependent enzyme links l-arginine to nitric oxide in N-nitroso biosynthesis. *J. Am. Chem. Soc.* 141, 4026–4033 (2019).
3. Zhao, G. et al. The biosynthetic gene cluster of pyrazomycin—a N-cysteine antibiotic with a rare pyrazole moiety. *ChemBioChem* 21, 644–649 (2020).
4. Hermanau, R. et al. Gramabactin is a bacterial siderophore with a diazenidindolide ligand system. *Nat. Chem. Biol.* 14, 841–843 (2018).
5. Chen, Y.-Y. et al. Molecular mechanism of azoxy bond formation for azoxycins biosynthesis. *Nat. Commun.* 10, 4420 (2019).
6. Zhao, G. et al. Nitric oxide as a source for bacterial triazole biosynthesis. *Nat. Commun.* 11, 1614 (2020).
7. Wang, B. & Cao, Z. Hydration of carbonyl groups: the labile H₃O⁺ ion as an intermediate modulated by the surrounding water molecules. *ChemBioChem* 20, 1145–1149 (2019).
8. Ren, D. et al. Identification of the C-glycoside synthases during biosynthesis of the pyrazole-C-nucleosides formycin and pyrazofurin. *Angew. Chem. Int. Ed.* 58, 16512–16516 (2019).
9. Wang, B. & Cao, Z. A novel formylating system: the l-arginine:N₃-forming enzyme constructs the pharmacophore of streptozotocin. *Phytochemistry* 67, 5–17 (2004).
10. Ibba, M. & Soll, D. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69, 617–650 (2000).
11. Garg, R. P., Qian, X. L., Alemany, L. B., Moran, S. & Parry, R. J. Investigations of valinomycin biosynthesis: elucidation of the role of seryl-tRNA. *Proc. Natl Acad. Sci. USA* 105, 6534–6543 (2008).
12. Wang, B. & Cao, Z. Hydration of carbonyl groups: the l-arginine:N₃-ion as an intermediate modulated by the surrounding water molecules. *Angew. Chem. Int. Ed. Engl.* 50, 3266–3270 (2011).
13. Zhou, H. et al. Chemo- and regioselective dihydroxylation of benzene to hydroquinone enabled by engineered cytochrome P450 monoxygenase. *Angew. Chem. Int. Ed. Engl.* 58, 764–768 (2019).
14. Waldman, A. J., Ng, T. L., Wang, P. & Balskus, E. P. Heteroatom bond formation in natural product biosynthesis. *Chem. Rev.* 117, 5784–5863 (2017).
15. Fersouzi, C., Majer, S. H., DiMucci, J. M. & Lancer, K. M. Biological and bioinspired inorganic N–N bond-forming reactions. *Chem. Rev.* 120, 5302–5320 (2020).
16. Del Rio Flores, A. et al. Biosynthesis of triacsin featuring an N-hydroxytriazene moiety. *ChemBioChem* 20, 1135–1139 (2019).
17. Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Press, 2001).
18. Richter, M., Marquetand, P., González-Vázquez, I., Sola, I. & González, L. SHARC: ab initio molecular dynamics with surface hopping in the adiabatic representation including arbitrary couplings. *J. Chem. Theory Comput.* 7, 1253–1258 (2011).
19. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multitithreading. *J. Comput. Chem.* 31, 455–461 (2010).
20. Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).
21. Maier, J. A. et al. f1f4SR: improving the accuracy of protein side chain and backbone parameters from f99SBF. *J. Chem. Theory Comput.* 11, 3696–3713 (2015).
22. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174 (2004).
23. Bayl, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280 (1993).
24. Case, D. A. et al. AMBER 2018 (University of California, 2018).
25. Frisch, M. J. et al. Gaussian 16, Revision B.01 (Gaussian, Inc., 2016).
26. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 113, 6378–6396 (2009).
27. Abdel-Azeim, S., Li, X., Chung, L. W. & Morokuma, K. Zinc–homeocysteine binding in cobalamin-dependent methionine synthase and its role in the
Acknowledgements

This work was supported by the National Natural Science Foundation of China (31872625 and 32122005) and the Zhejiang Provincial Natural Science Foundation (LR19C010001) to Y.-L.D., and the National Key R&D Program of China (2019YFA0606400) to B.W. We thank Zhiwei Ge (Analysis Center of Agrobiology and Environmental Sciences, Zhejiang University) for the assistance of LC-HR-MS/MS data collection.

Author contributions

G.Z. and Y.-L.D. carried out biochemical, bioinformatics and synthetic work, and performed structural elucidation. W.P. and B.W. performed computational studies on the reaction mechanism. K.S. and J.S. participated in the compound isolation. X.L. carried out NMR spectra collection on the isotope-labeled compounds. G.Z., W.P., B.W., and Y.-L.D. designed the study, analyzed the results, and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27523-x.

Correspondence and requests for materials should be addressed to Binju Wang or Yi-Ling Du.

Peer review information Nature Communications thanks Yohei Katsuyama and the other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021