Title

Discovery of the Azaserine Biosynthetic Pathway Uncovers a Biological Route for α-Diazoester Production

Journal

Angewandte Chemie, 135(28)

ISSN

0044-8249

Authors

Van Cura, Devon

Ng, Tai L

Huang, Jing

et al.

Publication Date

2023-07-10

DOI

10.1002/ange.202304646

Discovery of the Azaserine Biosynthetic Pathway Uncovers a Biological Route for α-Diazoester Production

Devon Van Cura[a], Tai L. Ng[a], Jing Huang[c][d], Harry Hager[a], John F. Hartwig [h], Jay D. Keasling[c][d][e][f][g], Emily P. Balskus*[a][b]

[a] Devon Van Cura, Dr. Tai L. Ng, Harry Hager, Prof. Emily P. Balskus
Department of Chemistry and Chemical Biology
Harvard University
Cambridge, MA, USA.
E-mail: balskus@chemistry.harvard.edu

[b] Prof. Emily P. Balskus
Howard Hughes Medical Institute
Harvard University
Cambridge, MA, USA.

[c] Dr. Jing Huang, Prof. Jay D. Keasling
Biological Systems and Engineering Division
Lawrence Berkeley National Laboratory
Berkeley, CA, USA.

[d] Dr. Jing Huang, Prof. Jay D. Keasling
Joint BioEnergy Institute
Lawrence Berkeley National Laboratory
Emeryville, CA, USA.

[e] Prof. Jay D. Keasling
Department of Chemical and Biomolecular Engineering
University of California, Berkeley
Berkeley, CA, USA.

[f] Prof. Jay D. Keasling
Synthetic Biochemistry Center
Institute for Synthetic Biology, Shenzhen Institute for Advanced Technologies
Shenzhen, China.

[g] Prof. Jay D. Keasling
Center for Biosustainability
Danish Technical University
Lyngby, Denmark.

[h] Prof. John F. Hartwig
Department of Chemistry
University of California, Berkeley
Berkeley, CA, USA.

Abstract: Azaserine is a bacterial metabolite containing a biologically unusual and synthetically enabling α-diazoester functional group. Here, we report the discovery of the azaserine (*aza*) biosynthetic gene cluster from *Glycomyces harbinensis*. Discovery of related gene clusters reveals previously unappreciated azaserine producers, and heterologous expression of the *aza* gene cluster confirms its role in azaserine assembly. Notably, this gene cluster encodes homologs of hydrazonoacetic acid (HYAA)-producing enzymes, implicating HYAA in α-diazoester biosynthesis. Isotope feeding and biochemical experiments support this hypothesis. These discoveries indicate that a 2-electron oxidation of a hydrazonoacetyl intermediate is required for α-diazoester formation, constituting a distinct logic for diazo biosynthesis. Uncovering this biological route for α-diazoester synthesis now enables the production of a highly versatile carbene precursor in cells, facilitating approaches for engineering complete carbene-mediated biosynthetic transformations *in vivo*.

Nitrogen–nitrogen (N–N) bond-containing functional groups are prominent in small molecule drugs and enabling chemical reagents.[1] Though more rare, these structural motifs are also found in microbial natural products, including secondary metabolites containing diazo, hydrazone, hydrazine, hydrazide, azo, azine, azoxy, *N*-nitroso, and *N*-hydroxytriazene groups.[2] Recent efforts have begun to reveal the biosynthetic enzymes involved in assembling these functional groups, enabling both the discovery of new natural products through genome mining and the elucidation of enzymatic strategies for accessing N–N bond-containing molecules.[3]

Diazo-containing molecules are particularly interesting due to their synthetic utility and application in bioorthogonal transformations.[4] Diazo compounds are powerful enabling reagents for C–H and X–H insertion reactions (X = N, O, S, P, Si), cyclopropanations, alkylations, ring expansions, and dipolar cycloadditions.[1b, 5] In particular, α-diazoesters, such as ethyldiazoacetate, have recently seen extensive use as substrates for engineered heme-dependent enzymes that perform stereoselective carbene-mediated reactions.[6] Diazo compounds are also used in bioorthogonal cycloaddition reactions with strained alkynes, offering improved yields and kinetics compared to analogous azide coupling partners under certain conditions.[7]

Though relatively few diazo-containing natural products have been reported (Fig. 1a), a large proportion of these compounds exhibit cytotoxic activity.[4] For example, the bioactivity of the diazofluorene antitumor antibiotics kinamycin A–D and lomaiviticin A derives from DNA damage inflicted by the reactive diazo warhead(s).[8] Little is known about the biosynthesis of diazo groups, with only four diazo-forming enzymes discovered to date. CreM, Aha11, and AvaA6 are related ATP-dependent ligases that catalyze diazotization of aromatic amines with nitrite during the biosynthesis of cremeomycin, tasikamides, and avenalumic acid, respectively.[9] AzpL, a membrane-bound enzyme unrelated to any characterized proteins, converts nitrite and 5-oxolysine to 6-diazo-5-oxonorleucine (DON) during the biosynthesis of the diazoketone natural product alazopeptin.[3a] In all of these cases, the biosynthetic strategy for diazo formation involves the production of nitrite by the L-aspartate-nitro-succinate (ANS) pathway[10] followed by an ATP-dependent reaction of nitrite with a nucleophilic primary amine (Figures 1b and S1). By contrast, the kinamycin biosynthetic pathway uses a distinct strategy for diazo formation involving early-stage production of a hydrazine synthon from nitrite that is transferred at a late stage to the natural product scaffold before oxidation to the diazo, however the diazo-forming enzyme in this pathway remains elusive (Figure S2).[11]

Discovered in 1954 from the soil microbe *Streptomyces fragilis* NRRL 2424,[12] azaserine is the first diazo-containing metabolite identified and one of only three known α-diazoester-containing natural products (Figure 1a).[4] As a cytotoxic glutamine antimetabolite, azaserine is used extensively to study glutamine dependent metabolic pathways. We became interested in the biosynthesis of azaserine because of its rare α-diazoester functionality and close structural resemblance to the important synthetic reagent ethyl diazoacetate. While a few diazo-containing metabolites have been accessed by heterologous

expression (diazoquinones and diazoketones), these molecules possess undesirable qualities for biological carbene transfer including instability and unsuitable reactivity.[3a, 9c, 13] As α-diazoesters are particularly important reagents for enzymatic carbene-mediated transformations, biological production of azaserine has great potential for biocatalytic and metabolic engineering applications, as we recently demonstrated in a related manuscript.[14]

Here, we report the discovery of the biosynthetic gene cluster that produces azaserine. Heterologous expression of the *aza* gene cluster and *in vitro* biochemical reconstitution of several key biosynthetic enzymes confirm the link between these genes and azaserine production. Isotope feeding and biochemical experiments, along with bioinformatic analysis of the *aza* gene cluster, implicate a hydrazone intermediate and highlight a distinct biosynthetic strategy for α-diazoester formation involving oxidation and incorporation of hydrazonoacetic acid (HYAA) (Figure 1b).

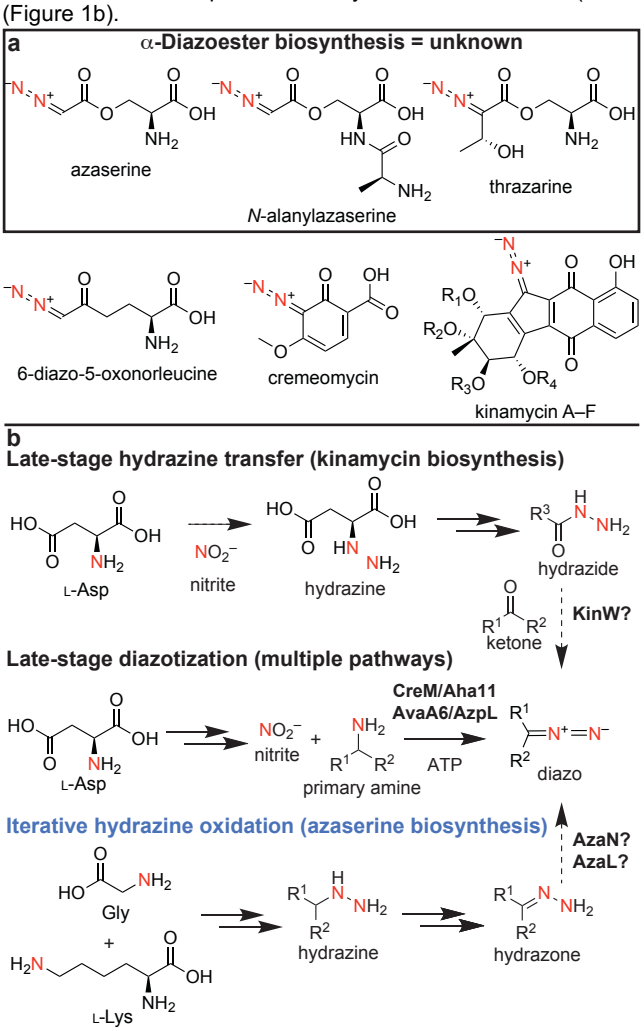


Figure 1. Microbes produce diazo-containing natural products using multiple biosynthetic strategies. **a**) Structures of selected diazo-containing natural products.[4] Azaserine, *N*-alanylazaserine, and thrazarine are the only reported α-diazoester-containing natural products. **b**) Comparison of the unique diazo formation strategy revealed in this study to previously reported pathways for diazo biosynthesis.

We initially sought to identify the azaserine biosynthetic gene cluster using comparative genomics. In addition to *S. fragilis*, the soil microbe *Glycomyces harbinensis* ATCC 43155 produces azaserine, and we confirmed this activity (Figure S3).[15] To identify putative azaserine biosynthetic genes, we sequenced the genomes of both organisms and performed BLAST searches using the known diazo-forming enzymes CreM, Aha11, AvaA6, and AzpL as queries, but we found no homologs conserved between both organisms.[9] Notably, BLAST searches with the ANS pathway enzymes, CreD and CreE, revealed homologs in *S. fragilis* but not *G. harbinensis*, suggesting that the azaserine pathway might employ a different biosynthetic logic for diazo assembly.

We next compared the *S. fragilis* and *G. harbinensis* genomes using the Mauve Genome Alignment Tool to identify conserved regions.[16] This analysis identified a ~25 kb gene cluster (*aza*) consisting of 19 open reading frames conserved in both organisms. Notably this was the only biosynthetic gene cluster conserved between the two organisms as predicted by antiSMASH.[17] The *aza* gene cluster contains a set of nine genes *azaABCEFGMOQ* that is homologous to s56-p1 and triacsin biosynthetic genes involved in converting L-lysine and glycine precursors to a carrier protein-tethered HYAA intermediate (Figures 2d and S4). The presence of genes encoding N–N bond forming enzymes in this shared gene cluster highlighted it as a promising candidate for azaserine biosynthesis.

To provide initial support for this assignment, we used the EFI Genome Neighborhood Tool[18] to identify related gene clusters in other sequenced bacterial genomes. This analysis revealed highly similar gene clusters in 46 additional strains from diverse phyla, including plant, animal, and human pathogens (Table S3 and Figure S5). LC-MS analysis of culture supernatants from two of these organisms, *Streptomyces ochraceiscleroticus* DSM 43155 and *Streptomyces olivoverticillatus* DSM 40250, revealed azaserine production, strengthening the functional assignment of the *aza* gene cluster (Figure 2a–b).

To unambiguously confirm this assignment, we sought to express the *aza* gene cluster in a heterologous host. Briefly, the gene cluster was excised from high molecular weight *G. harbinensis* genomic DNA using CRISPR/Cas9 with guide RNAs targeting either end of the cluster. A linearized dual-inducible integrative expression vector was prepared by PCR amplification with primers designed to introduce homology arms specific to the *aza* gene cluster. DNA fragments were joined by Gibson assembly to form the *aza*-pDualP expression vector and the sequence was verified by Sanger sequencing and restriction digest (Figure S6). *Aza*-pDualP was transferred to *Streptomyces albus* J1074, *Streptomyces coelicolor* M1152, and *Streptomyces lividans* TK64 via conjugation. Expression strains were cultured in a variety of growth media supplemented with ε-caprolactam and oxytetracycline inducers and assayed for azaserine production using LC-MS. All three heterologous hosts produced azaserine in the presence of *aza*-pDualP (Figure 2c), with *S. coelicolor* yielding the most robust production in the widest variety of media (Figure S7). These results definitively link the *aza* gene cluster to azaserine biosynthesis.

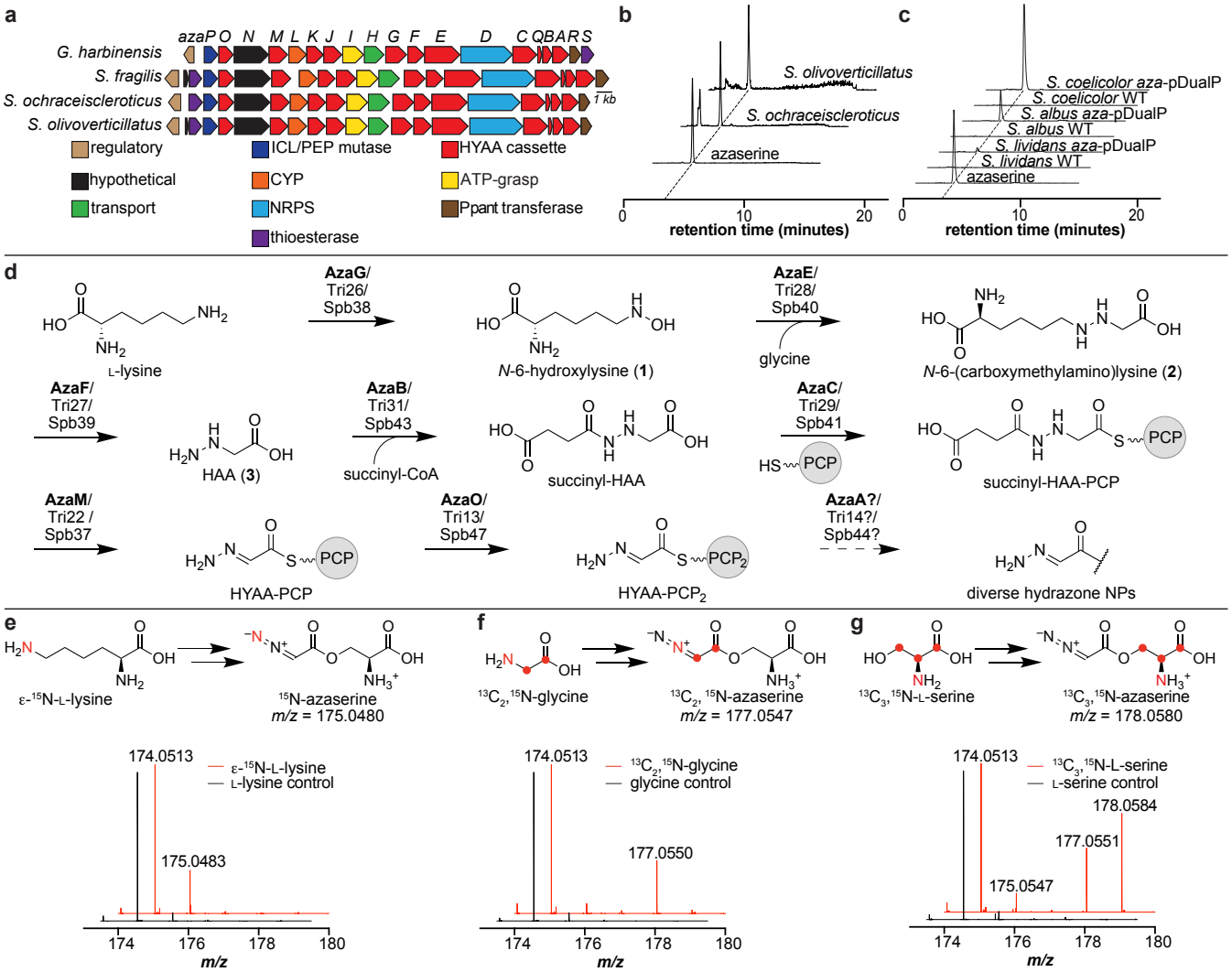


Figure 2. *In vivo* experiments link the *aza* gene cluster to azaserine production and suggest that azaserine biosynthesis proceeds via an HYAA intermediate. **a**) Azaserine-producing organisms encode a conserved gene cluster containing homologs of the HYAA-cassette identified in s56-p1 and triacsin biosynthesis. ICL/PEP = isocitrate lyase/phosphoenolpyruvate, CYP = cytochrome P450, NRPS = non-ribosomal peptide synthetase, Ppant = phosphopantetheine. **b**) Extracted ion chromatogram (EIC) for azaserine (*m/z* = 174.0509 ± 10 ppm) from extracts of two newly identified azaserine-producing organisms, *S. olivoverticillatus* and *S. ochraceiscleroticus*. Experiments were performed in biological duplicate and representative results are shown. **c**) EIC (*m/z* = 174.0509 ± 10 ppm) of extracts of wild-type *S. albus*, *S. coelicolor*, and *S. lividans* compared to the corresponding heterologous hosts transformed with *aza*-pDualP. Experiments were performed in biological triplicate and representative results are shown. **d)** Homologs of the HYAA biosynthetic enzymes from the triacsin, s56-p1, and azaserine pathways and their proposed biochemical transformations. PCP=Tri30, Spb42, and AzaQ; PCP_2=Tri20. There are no Tri20 homologs in the azaserine or s56-p1 biosynthetic gene clusters. **e**) Mass spectrum of ^{15}N-label incorporation into azaserine for *S. coelicolor aza*-pDualP cultures fed ε-^{15}N-L-lysine. Isotopically labelled atoms are shown in red. **f**) Mass spectrum of $^{13}C_2$,^{15}N-label incorporation into azaserine for *S. coelicolor aza*-pDualP cultures fed $^{13}C_2$,^{15}N-glycine. **g**) Mass spectrum of $^{13}C_3$,^{15}N-label incorporation into azaserine for *S. coelicolor aza*-pDualP cultures fed $^{13}C_3$,^{15}N-L-serine. Labeling experiments were performed in biological triplicate and representative results are shown.

We next analyzed the contents of the *aza* gene cluster to formulate a biosynthetic hypothesis and guide experimental characterization of the pathway. As highlighted earlier, the contents of the gene cluster suggested that azaserine biosynthesis would involve an HYAA intermediate. Conversion of this intermediate to azaserine would require subsequent condensation of the carboxylic acid of HYAA with the side-chain hydroxyl group of serine and oxidization of the hydrazone to the diazo group. The general biosynthetic strategy of generating a common N–N bond-containing intermediate that is shuttled into diverse biosynthetic pathways is also observed in the kinamycin and fosfazinomycin pathways.[11] Notably, this logic differs from other characterized pathways for diazo formation (Figures S1 and S2).

To confirm the involvement of an HYAA intermediate in azaserine biosynthesis, we performed stable isotope feeding experiments in *G. harbinensis*. Feeding ε-^{15}N-L-lysine to *G. harbinensis* cultures resulted in specific incorporation of a single ^{15}N atom into the diazo group of azaserine, consistent with production of an HYAA intermediate (Figure S8). Results of feeding ^{15}N-glycine were inconclusive, likely due to scrambling of the label by primary metabolism. To address this, we attempted feeding experiments with the heterologous expression strain *S. coelicolor aza*-pDualP, reasoning that induction of the *aza* gene cluster and addition of labeled amino acids during stationary phase might decrease scrambling. Feeding ε-^{15}N-L-lysine and $^{13}C_2$,^{15}N-glycine resulted in enrichment of ^{15}N- and ^{13}C,^{15}N-azaserine, respectively, compared to cultures fed unlabeled amino acids (Figure 2e-f). In both cases, MS/MS fragmentation demonstrated specific incorporation of labeled atoms into the diazoacetyl fragment of

azaserine (Figure S9). This suggests that the diazoacetyl group is derived from the C–N skeleton of glycine and the ε-N of lysine. While the observed MS/MS fragments do not directly distinguish between ^{15}N incorporation into the proximal or distal diazo nitrogen atoms, specific incorporation of intact $^{13}C_2$,^{15}N-glycine into the diazoacetyl fragment suggests that the proximal diazo nitrogen originates from glycine and the distal diazo nitrogen originates from L-lysine. These data are consistent with production of an HYAA intermediate during azaserine biosynthesis. Feeding $^{13}C_3$,^{15}N-L-serine to *S. coelicolor aza*-pDUalP resulted in enrichment of $^{13}C_3$, ^{15}N-azaserine (Figure 2g), suggesting that the remainder of the azaserine structure originates from serine. MS/MS fragmentation of $^{13}C_3$,^{15}N-azaserine (*m/z* = 178.0580) revealed specific incorporation of labels into the dehydroalanine fragment of azaserine (Figure S10). A significant amount of $^{13}C_2$,^{15}N-azaserine (*m/z* = 177.0551) was also observed in the same experiment (Figure 2g), and its MS/MS fragmentation was identical to that of labeled azaserine in cultures fed $^{13}C_2$,^{15}N-glycine (Figure S10). This suggests $^{13}C_2$,^{15}N-azaserine in this experiment arises from metabolism of $^{13}C_3$,^{15}N-serine to $^{13}C_2$,^{15}N-glycine before eventual incorporation into the diazoacetyl group.

To further probe the origin of the azaserine diazoacetyl group, we sought to reconstitute the activity of predicted HYAA biosynthetic enzymes (Figure 2d). As expected, lysine monooxygenase AzaG catalyzed the oxidation of lysine to *N*-6-hydroxylysine (**1**) (Figures S11-13). *E. coli* cultures expressing the N–N bond-forming MetRS/cupin enzyme AzaE linked **1** with glycine to generate *N*-6-(carboxymethylamino)lysine (**2**) (Figure S14). Attempts to express the D-amino acid oxidase-like enzyme AzaF, which is predicted to cleave a C–N bond of **2** to produce HAA (**3**), resulted in insoluble protein. Therefore, the putative product of this enzyme, **3**, was accessed synthetically to use as a substrate in downstream biochemical reactions.

We hypothesized that **3** would be converted to HYAA in a process analogous to triacsin biosynthesis (Figures 2d and S2). Consistent with this proposal, we demonstrated that HAA is succinylated by AzaB (Figure S15), though we also observed some non-enzymatic production of **4** at a reduced rate. An *in vitro* assay containing AzaBCMQ (homologs of Tri31, Tri29, Tri22, and Tri30, respectively), HAA, succinyl-CoA, ATP, and FAD yielded **6** as identified using a LC-MS-based Ppant ejection assay (Figure 3a). Together, these results provide biochemical support for involvement of an HYAA intermediate in azaserine biosynthesis.

AzaJ and AzaK are annotated as saccharopine dehydrogenases and homologs of these enzymes are often, but not always, co-localized with the HYAA cassette. Knockout of AzaJ and AzaK homologs in the triacsin pathway did not abolish triacsin production,[19] so we reasoned that these enzymes are likely not required for azaserine biosynthesis. Instead, we hypothesize that AzaJ and AzaK are involved in lysine salvage, as has been previously proposed in pyrazofurin biosynthesis.[20]

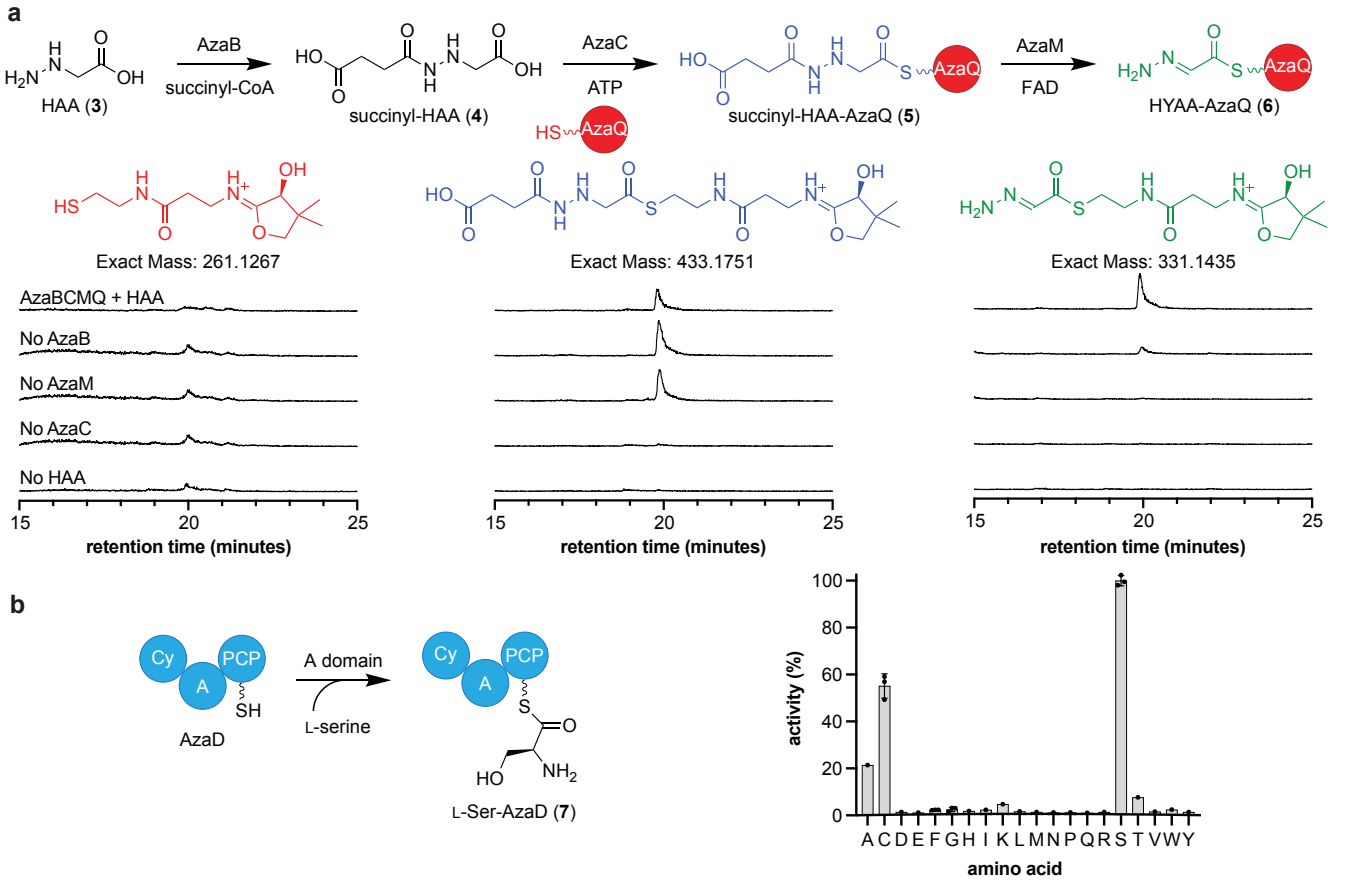


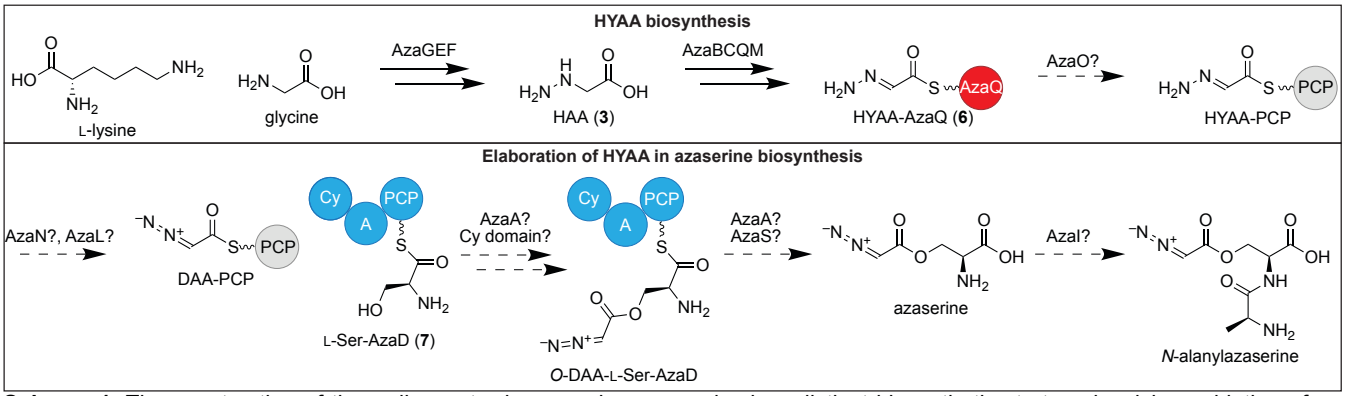
Figure 3. Biochemical assays suggest that the core structure of azaserine is derived from the condensation of HYAA and L-serine **a**) LC-MS Ppant ejection assay of the *in vitro* cascade reaction of AzaBCMQ. EICs of Ppant fragments are shown. The HYAA-Ppant fragment (*m/z* = 331.1435) was detected in the presence of HAA, succinyl-CoA, ATP, FAD, and AzaBCMQ. The succinyl-HAA-Ppant fragment (*m/z* = 433.1751) and HYAA-Ppant fragment were observed in the no AzaB control reaction due to non-enzymatic production of **4** under these conditions (Figure S15). In the absence of AzaM, accumulation of the succinyl-HAA-Ppant fragment was detected. The HAA-Ppant fragment (*m/z* = 333.1591) was not detected. Three biological replicates were performed. **b**) ATP-$^{32}PP_i$ exchange assay to assess substrate preference for the adenylation domain of AzaD. All twenty amino acids were tested in a single replicate and L-serine, L-cysteine, glycine, and L-phenylalanine were repeated in triplicate. Error bars represent mean ± std dev.

At this point, we hypothesized that azaserine biosynthesis would diverge from known HYAA-producing pathways. In the triacsin pathway, HYAA is transferred from peptidyl carrier protein (PCP) Tri30 onto another PCP (Tri20) by the transacylase Tri13 before being incorporated into a PKS-derived carbon scaffold and hydrolytically released from Tri20 by the C45-peptidase Tri14 (Figure S2). Further elaboration affords triacsins A–D. Since no Tri20 homologs are encoded in the azaserine gene cluster, we hypothesize that AzaO (Tri13 homolog) may transfer HYAA onto a yet unknown PCP. Intriguingly, comparison of the predicted structures of AzaE, Tri28, and Spb40 revealed an additional C-terminal domain found only in AzaE (Figure S16a–c) that resembles the predicted structure of Tri20, although it bears no discernable sequence homology. Its closest structural homolog is the carrier protein domain of carboxylic acid reductase (Figure S16d–f). Accordingly, we hypothesize the putative carrier protein domain of AzaE acts in place of Tri20 in azaserine biosynthesis. However, we cannot rule out the involvement of a PCP encoded elsewhere in the genome or the possibility that an additional PCP is not required for azaserine biosynthesis.

The AzaA homolog Tri14 has been reported to catalyze hydrolytic release of a PCP-bound intermediate in the triacsin pathway. We envision two potential functions for AzaA in azaserine production. C45 peptidases typically perform autoproteolytic cleavage to release a mature acyl transferase. The predicted structure of AzaA is most similar to the prototypical C45 peptidase acyl coenzyme A: 6-aminopenicillanic acid acyltransferase (Z-score = 39.5; Figure S17) which catalyzes *N*-transacylation during the biosynthesis of isopenicillin G by sequential amide hydrolysis and *N*-acyl transfer.[21] In azaserine biosynthesis, we envision that acyl transfer of HYAA from a carrier protein to the amine of serine could potentially be catalyzed by either AzaA or the Cy domain of AzaD before eventual isomerization to the ester by the Cy domain. Alternatively, the ester linkage could be generated directly by one of these enzymes. It is also possible AzaA catalyzes thioester hydrolysis of an AzaD-tethered intermediate, similar to Tri14.

The remaining genes in the *aza* gene cluster (*azaDHILNP*) encode proteins distinct from those in other HYAA-utilizing pathways. We hypothesize that a subset of these proteins transfer HYAA onto the serine side chain hydroxyl group and oxidize the hydrazonoacetyl group by two electrons en route to the final α-diazoester natural product. We propose that the non-ribosomal peptide synthetase (NRPS) AzaD likely plays a central role in linking HYAA to serine. This enzyme is predicted to contain an N-terminal TubC-like docking domain (TubCdd), a cyclization domain (Cy domain), an adenylation domain (A domain), and a peptidyl carrier protein domain (PCP domain).[22] Bioinformatic predictions initially suggested this A domain would activate cysteine, however, ATP-^{32}PP$_i$ exchange assays with full-length AzaD revealed preferential activation of serine (Figure 3b). This supports a role for AzaD in linking HYAA or diazoacetic acid (DAA) to serine, depending on the timing of diazo formation. The presence of a Cy domain in AzaD is particularly intriguing. This domain typically catalyzes amide bond formation followed by attack of a Ser side chain hydroxyl or a Cys side chain thiol onto the amide carbonyl to generate an oxazoline or thiazoline heterocycle.[23] Because the formation of azaserine necessitates a similar nucleophilic attack of a serine side chain hydroxyl group onto an HYAA- or DAA-based electrophile, we hypothesize that the Cy domain of AzaD may mediate this transformation (Scheme 1). Predicted thioesterase AzaS may hydrolytically release intermediates from the PCP domain of AzaD and/or other carrier protein-tethered intermediates. AzaI is predicted to be a member of the functionally diverse ATP-grasp superfamily. Dali searches using the predicted structure of AzaI revealed its closest characterized structural homolog is an L-amino acid ligase RizA (4wd3, Z-score = 32.1).[24] The predicted structure of AzaI also bears striking resemblance to the predicted structure of AzpA which was recently reported to form an *N*-alanyl peptide during the biosynthesis of the α-diazoketone natural product alazopeptin (Figure S18).[3a] Accordingly, we hypothesize that AzaI may be involved in peptide bond formation to produce *N*-alanylazaserine which has been detected from *G. harbinensis* grown in minimal media in previous studies.[15] Interestingly, *N*-alanylazaserine was not observed from *S. coelicolor aza*-pDualP during growth in complex or minimal media, implying some yet unknown regulatory mechanism for *N*-alanylazaserine production. It is also possible that *N*-alanylazaserine is formed as a shunt product under certain growth conditions, or that it is produced by genes outside of the azaserine gene cluster.

Generating the key α-diazoester functional group from a hydrazone intermediate requires oxidative chemistry. AzaL and AzaN are the only predicted oxidoreductases in the azaserine pathway without an assigned role, making them the most likely candidates for diazo formation. AzaL resembles cytochrome P450 enzymes which catalyze heme-dependent oxidation and oxygenation reactions. The closest predicted structural homologs of AzaL are CalO2 and P450$_{BioI}$ (Figure S19). Though its activity has not been experimentally verified, preliminary substrate binding studies suggest CalO2 likely hydroxylates an ACP- or CoA-tethered substrate.[25] P450$_{BioI}$ catalyzes multiple hydroxylations of an ACP-tethered substrate, leading to oxidative C–C bond cleavage.[26] The predicted structural similarity of AzaL to these enzymes suggested an ability to oxidize a carrier protein-tethered intermediate, potentially derived from HYAA, during azaserine biosynthesis. Though annotated as a hypothetical protein, Conserved Domain Database[27] searches suggest AzaN resembles heme-oxygenase-like diiron enzymes (HDOs) which typically harbor a diiron metallocofactor and perform oxidation/oxygenation chemistry.[28] Interestingly, the characterized protein that best resembles the predicted structure of AzaN is SznF, which performs *N*-oxygenation and N–N bond formation in streptozotocin biosynthesis (Figure S20).[28a] Further experimental work is needed to understand the roles of these two oxidoreductases in azaserine biosynthesis.



Scheme 1. The construction of the α-diazoester in azaserine proceeds via a distinct biosynthetic strategy involving oxidation of an HYAA intermediate. The activities of HYAA-forming enzymes are shown in the top box. Proposed elaboration of the HYAA intermediate in azaserine biosynthesis is shown in the lower box.

In addition to the identity of the diazo forming enzyme(s), the timing of this transformation in azaserine biosynthesis remains unclear. In the triacsin pathway, HYAA is produced by enzymes encoded in the HYAA cassette and used as a synthon for convergent biosynthesis with a PKS-derived scaffold. A similar convergent biosynthetic strategy is likely involved in the biosynthesis of the structurally dissimilar natural product, s56-p1.[29] However, only the production of HAA has been verified in the s56-p1 pathway (Figure S2b). While gene clusters encoding putative diazo biosynthetic enzymes almost always encode the HYAA cassette, they frequently lack homologs of the NRPS AzaD. Notably, the gene clusters missing AzaD encode assembly line enzymes predicted to form distinct intermediates (Figure S21). This implies that HYAA (or DAA) may be linked to scaffolds other than serine in these pathways. The potential diversity of natural product structures produced by these additional gene clusters may suggest that diazo formation occurs on the conserved HYAA synthon prior to transfer onto structurally diverse natural product scaffolds. However, we cannot exclude the possibility that diazo formation occurs later in the azaserine pathway. Taken together, these bioinformatic analyses highlight AzaL and AzaN as candidate diazo-forming enzymes, suggest a plausible biosynthetic route to azaserine from HYAA-AzaQ and L-serine-AzaD intermediates, and reveal the potential for biosynthesis of diverse diazo-containing metabolites.

In summary, we have identified the biosynthetic pathway that produces the biologically and synthetically important α-diazoester natural product azaserine. Bioinformatic analyses, isotope feeding experiments, and *in vitro* biochemical experiments indicate that the logic of diazo formation in this pathway, early-stage generation of an HYAA synthon, is distinct from previously known strategies for diazo biosynthesis. This work sets the stage for investigation of the diazo-forming enzymes in this pathway which will further the nascent understanding of the biosynthesis of diazo natural products. Further, the identification of closely related gene clusters in human, plant, and animal pathogens suggests potentially interesting ecological roles for diazo metabolites. Finally, genetically encoded α-diazoester production will enable multiple biocatalytic, metabolic engineering, and biorthogonal chemistry applications including the construction of complete metabolic pathways for carbene-mediated chemistry *in vivo* using engineered heme-dependent enzymes, as we demonstrate in a separate study.[14]

Author Contributions

D.V.C., T.L.N., and E.P.B. conceived the project. D.V.C. and E.P.B. wrote the manuscript with input from all other authors. D.V.C., T.L.N, J.D.K., and J.H. identified the azaserine gene cluster. T.L.N. prepared samples for genome sequencing. D.V.C. and H.H. expressed and purified proteins. D.V.C. performed the remaining experiments.

Acknowledgements

We thank J. Wang for assistance with LC-MS data collection. D.V.C. acknowledges funding from the National Science Foundation Graduate Research Fellowship Program (grant numbers DGE2140743 and DGE1745303). We acknowledge financial support from the National Institute of Health (grant numbers 1DP2GM105434-01 and 5R01GM132564-03), the Joint BioEnergy Institute (https://www.jbei.org), which is supported by the DOE, Office of Science, Office of Biological and Environmental Research under contract DE-AC02-05CH11231, and the National Science Foundation (grant number 2027943). E.P.B is a Howard Hughes Medical Institute Investigator. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Competing interests: J.D.K. has a financial interest in Amyris, Demetrix, Maple Bio, Lygos, Napigen, Berkeley Yeast, Zero Acre Farms, Ansa Biotechnologies, Apertor Pharmaceuticals, ResVit Bio, and Cyklos Materials. The other authors declare no competing interests.

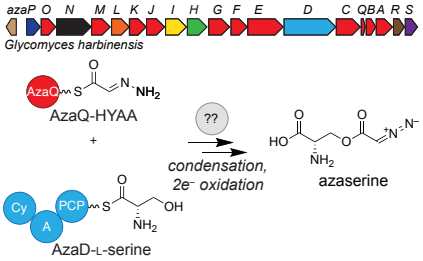
Keywords: azaserine • biosynthesis • diazoesters • N–N bonds • natural products

References

[1] a) E. A. Ilardi, E. Vitaku, J. T. Njardarson, *J. Chem. Ed.* **2013**, *90*, 1403–1405; b) A. Ford, H. Miel, A. Ring, C. N. Slattery, A. R. Maguire, M. A. McKervey, *Chem. Rev.* **2015**, *115*, 9981–10080.
[2] L. M. Blair, J. Sperry, *J. Nat. Prod.* **2013**, *76*, 794–812.
[3] a) S. Kawai, Y. Sugaya, R. Hagihara, H. Tomita, Y. Katsuyama, Y. Ohnishi, *Angew. Chem. Int. Ed.* **2021**, *60*, 10319–10325; b) L. Chen, Z. Deng, C. Zhao, *ACS. Chem. Biol.* **2021**, *16*, 559–570; c) H. Y. He, H. Niikura, Y. L. Du, K. S. Ryan, *Chem. Soc. Rev.* **2022**, *51*, 2991–3046.
[4] C. C. Nawrat, C. J. Moody, *Nat. Prod. Rep.* **2011**, *28*, 1426–1444.
[5] K. A. Mix, M. R. Aronoff, R. T. Raines, *ACS. Chem. Biol.* **2016**, *11*, 3233–3244.
[6] Y. Yang, F. H. Arnold, *Acc. Chem. Res.* **2021**, *54*, 1209–1225.
[7] a) K. A. Andersen, M. R. Aronoff, N. A. McGrath, R. T. Raines, *J. Am. Chem. Soc.* **2015**, *137*, 2412–2415; b) N. A. McGrath, R. T. Raines, *Chem. Sci.* **2012**, *3*, 3237–3240.
[8] a) C. M. Woo, Z. Li, E. K. Paulson, S. B. Herzon, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2851–2856; b) S. B. Herzon, *Acc. Chem. Res.* **2017**, *50*, 2577–2588; c) S. B. Herzon, C. M. Woo, *Nat. Prod. Rep.* **2012**, *29*, 87–118.
[9] a) A. J. Waldman, E. P. Balskus, *J. Org. Chem.* **2018**, *83*, 7539–7546; b) G. L. Ma, H. Candra, L. M. Pang, J. Xiong, Y. Ding, H. T. Tran, Z. J. Low, H. Ye, M. Liu, J. Zheng, M. Fang, B. Cao, Z. X. Liang, *J. Am. Chem. Soc.* **2022**, *144*, 1622–1633; c) S. Kawai, R. Hagihara, K. Shin-Ya, Y. Katsuyama, Y. Ohnishi, *Angew. Chem. Int. Ed.* **2022**, *61*, e202211728.
[10] Y. Sugai, Y. Katsuyama, Y. Ohnishi, *Nat. Chem. Biol.* **2016**, *12*, 73–75.
[11] K. A. Wang, T. L. Ng, P. Wang, Z. Huang, E. P. Balskus, W. A. van der Donk, *Nat. Commun.* **2018**, *9*, 3687.
[12] Q. R. Bartz, C. C. Elder, R. P. Frohardt, S. A. Fusari, T. H. Haskell, D. W. Johannessen, A. Ryder, *Nature* **1954**, *173*, 72–73.
[13] a) X. Liu, D. Liu, M. Xu, M. Tao, L. Bai, Z. Deng, B. A. Pfeifer, M. Jiang, *J. Nat. Prod.* **2018**, *81*, 72–77; b) A. J. Waldman, Y. Pechersky, P. Wang, J. X. Wang, E. P. Balskus, *Chembiochem* **2015**, *16*, 2172–2175.
[14] J. Huang, A. Quest, P. Cruz-Morales, K. Deng, J. H. Pereira, D. Van Cura, R. Kakumanu, E. E. K. Baidoo, Q. Dan, Y. Chen, C. J. Petzold, T. R. Northen, P. D. Adams, D. S. Clark, E. P. Balskus, J. F. Hartwig, A. Mukhopadhyay, J. D. Keasling, *Nature* **2023**. https://doi.org/10.1038/s41586-023-06027-2
[15] M. D. Lee, A. A. Fantini, N. A. Kuck, M. Greenstein, R. T. Testa, D. B. Borders, *J. Antibiot.* **1987**, *40*, 1657–1663.
[16] A. C. Darling, B. Mau, F. R. Blattner, N. T. Perna, *Genome Res.* **2004**, *14*, 1394–1403.
[17] K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2021**, *49*, W29–W35.
[18] R. Zallot, N. Oberg, J. A. Gerlt, *Biochemistry* **2019**, *58*, 4169–4182.
[19] A. Del Rio Flores, F. F. Twigg, Y. Du, W. Cai, D. Q. Aguirre, M. Sato, M. J. Dror, M. Narayanamoorthy, J. Geng, N. A. Zill, R. Zhai, W. Zhang, *Nat. Chem. Biol.* **2021**, *17*, 1305–1313.
[20] M. Zhang, P. Zhang, G. Xu, W. Zhou, Y. Gao, R. Gong, Y. S. Cai, H. Cong, Z. Deng, N. P. J. Price, X. Mao, W. Chen, *Appl. Environ. Microbiol.* **2020**, *86*, e01971-19.
[21] M. Bokhove, H. Yoshida, C. M. Hensgens, J. M. van der Laan, J. D. Sutherland, B. W. Dijkstra, *Structure* **2010**, *18*, 301–308.
[22] B. O. Bachmann, J. Ravel, *Methods Enzymol.* **2009**, *458*, 181–217.
[23] a) K. Bloudoff, C. D. Fage, M. A. Marahiel, T. M. Schmeing, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 95–100; b) D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh, C. L. Drennan, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12432–12437.
[24] W. Kagawa, T. Arai, S. Ishikura, K. Kino, H. Kurumizaka, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2015**, *71*, 1125–1130.
[25] J. G. McCoy, H. D. Johnson, S. Singh, C. A. Bingman, I. K. Lei, J. S. Thorson, G. N. Phillips, Jr., *Proteins: Struct., Funct., Bioinf.* **2009**, *74*, 50–60.

[26] M. J. Cryle, I. Schlichting, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15696–15701.
[27] a) S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, *Nucleic Acids Res.* **2020**, *48*, D265–D268; b) A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, L. Y. Geer, S. H. Bryant, *Nucleic Acids Res.* **2017**, *45*, D200–D203; c) A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, S. H. Bryant, *Nucleic Acids Res.* **2015**, *43*, D222–D226; d) A. Marchler-Bauer, S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S. H. Bryant, *Nucleic Acids Res.* **2011**, *39*, D225–D229; e) A. Marchler-Bauer, S. H. Bryant, *Nucleic Acids Res.* **2004**, *32*, W327–W331.
[28] a) T. L. Ng, R. Rohac, A. J. Mitchell, A. K. Boal, E. P. Balskus, *Nature* **2019**, *566*, 94–99; b) B. Zhang, L. J. Rajakovich, D. Van Cura, E. J. Blaesi, A. J. Mitchell, C. R. Tysoe, X. Zhu, B. R. Streit, Z. Rui, W. Zhang, A. K. Boal, C. Krebs, J. M. Bollinger, Jr., *J. Am. Chem. Soc.* **2019**, *141*, 14510–14514; c) L. Pang, W. Niu, Y. Duan, L. Huo, A. Li, J. Wu, Y. Zhang, X. Bian, G. Zhong, *Eng. Microbiol.* **2022**, *2*, 100007; d) J. B. Hedges, K. S. Ryan, *Angew. Chem. Int. Ed.* **2019**, *58*, 11647–11651; e) O. M. Manley, H. Tang, S. Xue, Y. Guo, W. C. Chang, T. M. Makris, *J. Am. Chem. Soc.* **2021**, *143*, 21416–21424; f) J. A. Marchand, M. E. Neugebauer, M. C. Ing, C. I. Lin, J. G. Pelton, M. C. Y. Chang, *Nature* **2019**, *567*, 420–424.
[29] K. Matsuda, T. Tomita, K. Shin-Ya, T. Wakimoto, T. Kuzuyama, M. Nishiyama, *J. Am. Chem. Soc.* **2018**, *140*, 9083–9086.

Entry for the Table of Contents



Discovery of the biosynthetic gene cluster for the α-diazoester natural product azaserine is reported. Isotope feeding and biochemical experiments implicate generation of a hydrazonoacetic acid intermediate that is oxidized and transferred to L-serine. This pathway represents a distinct biosynthetic strategy for diazo formation.

Institute and/or researcher Twitter usernames: @balskuslab, @devonvancura1, @harvardccb @MicrobialChem