PNAS RESEARCH ARTICLE | BIOCHEMISTRY

Multisubstrate specificity shaped the complex evolution of the aminotransferase family across the tree of life

Kaan Koper[a,1], Sang-Woo Han[b,c,1], Ramani Kothadia[d], Hugh Salamon[d,2], Yasuo Yoshikuni[b,d,e,f,g,h,3], and Hiroshi A. Maeda[a,3]

Affiliations are included on p. 10.

Edited by Clinton Chapple, Purdue University, West Lafayette, IN; received April 2, 2024; accepted May 14, 2024

Aminotransferases (ATs) are an ancient enzyme family that play central roles in core nitrogen metabolism, essential to all organisms. However, many of the AT enzyme functions remain poorly defined, limiting our fundamental understanding of the nitrogen metabolic networks that exist in different organisms. Here, we traced the deep evolutionary history of the AT family by analyzing AT enzymes from 90 species spanning the tree of life (ToL). We found that each organism has maintained a relatively small and constant number of ATs. Mapping the distribution of ATs across the ToL uncovered that many essential AT reactions are carried out by taxon-specific AT enzymes due to wide-spread nonorthologous gene displacements. This complex evolutionary history explains the difficulty of homology-based AT functional prediction. Biochemical characterization of diverse aromatic ATs further revealed their broad substrate specificity, unlike other core metabolic enzymes that evolved to catalyze specific reactions today. Interestingly, however, we found that these AT enzymes that diverged over billion years share common signatures of multisubstrate specificity by employing different nonconserved active site residues. These findings illustrate that AT family enzymes had leveraged their inherent substrate promiscuity to maintain a small yet distinct set of multifunctional AT enzymes in different taxa. This evolutionary history of versatile ATs likely contributed to the establishment of robust and diverse nitrogen metabolic networks that exist throughout the ToL. The study provides a critical foundation to systematically determine diverse AT functions and underlying nitrogen metabolic networks across the ToL.

enzyme family evolution | core metabolism | substrate promiscuity | nitrogen metabolism | multifunctional enzymes

Significance

The tree of life (ToL)-wide analyses of the ubiquitous aminotransferases (AT) family revealed that the broad substrate promiscuity of ATs, which is unusual for core metabolic enzymes, allowed recruitment of distinct, nonorthologous ATs to carry out essential AT reactions in different taxa but without increasing their copy numbers. Some distantly related ATs were also found to exhibit a common signature of multisubstrate specificity by employing different nonconserved active site residues. The versatile evolutionary trajectory of the promiscuous AT enzyme family likely led to biochemical diversity of the robust nitrogen metabolic networks that exist among various extant organisms.

It is hypothesized that the primordial enzymes were highly promiscuous and were able to catalyze a broad spectrum of related reactions. Although such catalysts were likely inefficient, a small set of these multifunctional enzymes could provide the necessary biochemical diversity to sustain ancient metabolism (1–3). Later, these primordial enzymes divergently evolved through gene duplication and specialization of each paralog to catalyze specific biochemical reactions, so as to increase metabolic efficiency without needing to increase overall protein expression (4, 5). The resulting rapid expansion of biochemical toolkits likely shaped the core metabolism of the last universal common ancestor (LUCA) (6–9).

The evolution of many ancient protein families involved in core metabolism has been studied across the tree of life (ToL). The Superfamily Classification of Protein database classifies proteins based on their structural and mechanistic similarities and articulates groups of modern enzyme families that likely originated from common ancestral enzymes, or founder enzymes (10). Some superfamilies—e.g., ribosomes (11), aminoacyl-tRNA synthetases (12–14), carbonic anhydrases (15), peptidases (16, 17)—are present across all extant organisms and hence were likely present in LUCA (18). Many of these LUCA enzymes also diverged, functionalized, and specialized to create complex metabolic networks across the ToL. These essential core metabolic enzymes were, in general, inherited vertically to descendants, with some occurrences of lateral gene transfers that replaced the functions of orthologous enzymes that derived from the same founder enzyme (19–21). Therefore, homology-based prediction of functional annotation is largely effective for most core metabolic enzymes.

Aminotransferases (ATs) are one notable exception and appear to retain high substrate promiscuity, posing significant challenges in the homology-based functional prediction of individual AT proteins or assigning specific AT activities to particular enzymes (22, 23). AT enzymes belong to pyridoxal 5′-phosphate (PLP)-dependent transferases superfamily (Fig. 1*A* and *SI Appendix*, Fig. S1*A* and Table S1) and play central roles in core

Author contributions: K.K., S.-W.H., Y.Y., and H.A.M. designed research; K.K., S.-W.H., R.K., H.S., and H.A.M. performed research; H.S. contributed new reagents/analytic tools; K.K., S.-W.H., and H.A.M. analyzed data; and K.K., S.-W.H., Y.Y., and H.A.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]K.K. and S.-W.H. contributed equally to this work.

[2]Present address: Knowledge Synthesis Inc., Berkeley, CA 94710.

[3]To whom correspondence may be addressed. Email: yyoshikuni@lbl.gov or maeda2@wisc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405524121/-/DCSupplemental.

Published June 17, 2024.

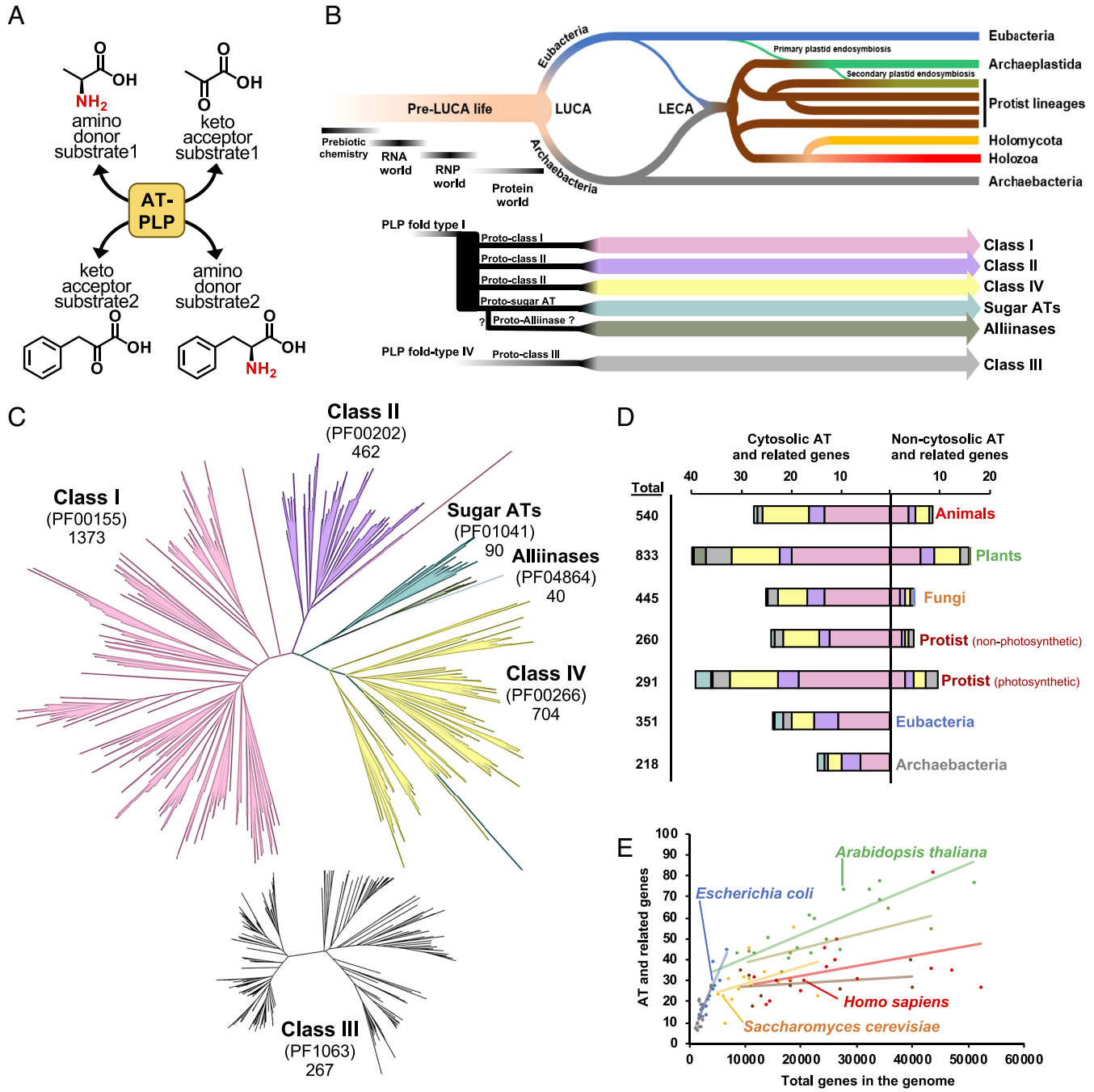

Fig. 1. AT enzymes evolved with a limited copy number expansion across the ToL. (*A*) ATs are PLP-dependent enzymes, which catalyze reversible transamination reactions among at least four substrates. (*B*) Evolution of transamination since the origin of life to today. Pre-LUCA life likely had nonenzymatic, RNA, or RNP-based transamination. At the interface of RNP and protein worlds, protein-based transaminases appeared in the form of proto-ATs classes. Proto-AT classes were inherited by LUCA and its descendants. (*C*) Phylogenetic analysis of AT candidate genes from seven Pfam domains known to contain ATs. Classes I, II, and IV, alliinase and sugar ATs are part of PLP-fold type I, while class III is a part of the independently evolved PLP-fold type IV. (*D*) Average numbers, Pfam domain composition, and subcellular localization of AT and related genes from animals, plants, fungi, protists, eubacteria, and archaea. (*E*) The relationship between the number of AT and related genes vs. total number of genes per species. Data corresponding to key model species are labeled, and the lines show the overall trend per taxon. Blue, eubacteria; gray, archaea; brown, nonphotosynthetic protists; army green, photosynthetic protists; orange, fungi; green, plants; red, animals.

nitrogen metabolism, essential to all organisms (23, 24). For instance, alanine (Ala) ATs from various organisms transaminate Ala to pyruvate for gluconeogenesis, thereby connecting central carbon and nitrogen metabolisms (23, 25). AT-mediated distribution of nitrogen is a key determinant of a wide range of critical biological processes, such as amino acid and protein homeostasis, synthesis of neurotransmitters, pathogenesis of infectious diseases, and recycling of nitrogen (23, 26, 27). Thus, deeper understanding of AT functions can lead to better disease diagnostics (e.g., Ala AT as liver damage marker), improved production of nitrogenous compounds (e.g., essential amino acids, alkaloid natural products) (28), and enhanced nitrogen use efficiency in crops (29). However, the functions of many AT enzymes remain poorly defined, leading to the limited understanding of nitrogen metabolic networks that operate in different organisms.

It is estimated that ancestral ATs emerged during the ribonucleoprotein (RNP) world and diversified before LUCA (23), and, today, are ubiquitously present across the ToL (Fig. 1*B*). Regardless

of their essentiality and billions of years of evolution, ATs still often retain a broad substrate specificity likely due to their inherit mechanistic constraint; ATs catalyze the reversible transamination reaction between amino donors and keto accepters using the ping-pong bi–bi mechanism in a single active site, requiring sufficient versatility to accommodate at least four different substrates. Also, the reactive PLP must be shielded from nonspecific substrates to minimize inhibitions (Fig. 1*A* and *SI Appendix*, Fig. S1*A*) (23, 30). Nonetheless, all extant organisms must maintain robust and complex nitrogen metabolic networks (28, 31), allowing for the distribution of an essential nutrient, nitrogen, to every corner of organismal metabolism, regardless of its availability in different environments. It is unknown, however, how AT enzymes have overcome these inherent mechanistic constraints associated with AT reactions and maintained the core nitrogen metabolism.

To address these questions, this study analyzed AT and related enzymes in 90 species across the ToL, which include bacteria, archaea, plants, fungi, animals, and various protist taxa (*SI Appendix*, Table S2 and Fig. S1*B*). The deep evolutionary analyses revealed several interesting paths that the AT enzyme family uniquely adapted. Despite the huge variations in genome sizes and complexities across the ToL, the number of ATs per species remains relatively small and constant. Many essential AT reactions are carried out by taxon-specific AT enzymes due to wide-spread instances of nonorthologous gene displacements (32) by distantly related AT groups. This finding explained why homology-based predictions of AT functions are difficult and highlighted the existence of missing enzymes catalyzing essential ATs in certain taxa. Some of these distantly related AT enzymes were further found to exhibit conserved multisubstrate specificity, which is accomplished by recruiting different nonconserved active site residues to recognize the same substrate. These results illuminate how AT family enzymes had overcome their inherent mechanistic constraints to maintain a relatively small but distinct set of AT enzymes in different taxa that likely support robust and diverse nitrogen metabolic networks that exist across the ToL.

Results

A Small Set of ATs Maintains Essential Core Nitrogen Metabolism across the ToL. In core metabolism, a single or a few enzymes, with high specificity and efficiency, are typically responsible for catalyzing each biochemical step. However, ATs often retain broad substrate specificity (23), which is typically seen in secondary or specialized metabolic enzymes that underwent significant copy number expansion (33–38). We therefore investigated how many ATs are present in various extant organisms by obtaining and analyzing all putative AT sequences from the genomes of 90 organisms across the ToL. We queried the genomes of 15 species from eubacteria (39), archaea (40), archaeplastida (plants) (41, 42), holozoa (animals) (43), holomycota (fungi) (44), as well as major protist taxa, spanning multiple phyla (45) (*SI Appendix*, Fig. S1*B* and Table S2), for protein sequences containing conserved protein family (Pfam) domains known for AT enzymes (*SI Appendix*, Tables S1 and S2). These species were selected based on the availability of high-quality genomes that represent the diversity of phyla and clades (*SI Appendix*, Fig. S1*B* and Table S2). Inspection of Pfam by HMMER (46) corrected the Pfam annotations for 93 ATs (Dataset S1) and a length criterion (250 to 1,200 amino acids) excluded potential pseudogenes, resulting in 2,938 putative AT genes (Dataset S2). Since our ability to detect homology fades over long evolutionary timescales (47, 48), a structure-assisted multiple sequence alignment, MAFFT-DASH (49), was used to capture the deep evolutionary kinship among distantly related ATs (*Materials and Methods*).

ATs are found among two independently evolved PLP-dependent enzyme fold-types, I and IV (Fig. 1*B*) (23, 24). The fold-type I includes canonical AT families of class I, II, and IV, as well as smaller families of sugar ATs and alliinases, whereas the fold-type IV only comprises class III ATs. We therefore separately aligned sequences from each Pfam family, having a close phylogenetic proximity (50), and then merged into two master alignments for independently evolved fold-type I and IV (Fig. 1*C*). The phylogeny estimation using the approximately maximum likelihood method (51) captured the monophyletic origins of AT and related enzymes from different Pfam families (Fig. 1*C*). Overall, class I (PF00155) represented the largest clade having 1,373 out of the 2,938 putative AT enzymes, followed by 704 in class IV (PF00266), 462 in class II (PF00202), and 267 in class III (PF01063, Fig. 1*C* and *SI Appendix*, Table S2). Ninety sugar ATs (PF01041), involved in bacterial cell wall biosynthesis (52, 53), were enriched in eubacteria and archaea, while 40 alliinases (PF04864) were mainly restricted to plants and protists with secondary plastids, except for two amorphean protists without secondary plastids (*Thecamonas trahens* and *Monosiga brevicollis*, *SI Appendix*, Table S2).

The copy number analyses showed that archaea and eubacteria on average have ~15 and ~25 AT and related genes, respectively, though a few prokaryotes having parasitic, pathogenic, or symbiotic lifestyles (54–56) had much lower numbers of ATs (Fig. 1*D* and *SI Appendix*, Table S2). In theory, up to 22 AT activities are needed to establish core nitrogen metabolism for synthesis of essential amino acids and nucleic acids (23). Also, genome reconstruction studies of LUCA (57) suggest that ~25 AT and related genes (from classes I, II, III, IV, and sugar ATs) were present at the base of the ToL. Therefore, these analyses revealed that the copy numbers of AT enzymes remained constant in prokaryotes.

Eukaryotes had a higher total number of AT and related genes than archaea and eubacteria (Fig. 1*D*, *SI Appendix*, Fig. S2 *A* and *B*, and Dataset S3), with plants and photosynthetic protists having the highest numbers (Fig. 1*D*, *SI Appendix*, Fig. S2 *A–C*, and Dataset S3). This likely reflects the expansion of metabolic processes [e.g., photorespiration (58), chlorophyll biosynthesis (59)] necessary to maintain an autotrophic lifestyle. Aside from these photosynthetic organisms, the higher AT number of eukaryotes was attributed to functionally homologous AT isoforms for organelle targeting (Fig. 1*D*): e.g., three mammalian aspartate (Asp) AT copies with highly redundant functions (60, 61). Additionally, multicellular eukaryotes had tissue-specific AT paralogs, such as human AlaAT1 and AlaAT2 (62) and Arabidopsis tryptophan (Trp) ATs, TAA1, TAR1, and TAR2 (63, 64). However, the average number of AT candidate genes was only 1.5 and 2.4 times higher in animals and plants, respectively, than in eubacteria, whereas the average number of genes in the genomes of animals and plants are 8.1 and 6.8 times larger than those in eubacteria (Fig. 1*E* and *SI Appendix*, Fig. S2*C* and Table S2). This indicates that, despite the substantial expansion of eukaryotic genome sizes and metabolic networks (65), AT gene families did not experience substantial copy number expansion even within eukaryotes. This is in contrast to the evolutionary history of specialized metabolic enzymes, many of which are also promiscuous but underwent significant copy number expansion (33, 34). These data revealed that organisms can maintain a robust and complex core nitrogen metabolic network with a relatively small number of AT genes.

Essential AT Reactions Are Redundantly Catalyzed by Distinct Sets of Distant, Nonorthologous ATs in Different Taxa. Ancestral ATs diversified into at least six different classes before LUCA and,

today, are ubiquitously present across the ToL (Fig. 1*B*). However, it is unknown how these founder AT enzymes of LUCA evolved to catalyze essential AT reactions of core nitrogen metabolism, hindering the accurate prediction of AT functions and nitrogen metabolic networks that exist in various organisms. To address this issue, based on the large-scale AT phylogeny illustrating an overall relationship of Pfam families (Fig. 1*C*), we further built 40 FastTree trees using multiple starting trees for each AT class to avoid a potential issue of local optima (66). We then identified 62 distinct AT and related enzyme groups that represent monophyletic clades (bootstrap value ≥ 0.9) (*SI Appendix*, Figs. S3–S7 and Dataset S3). The "AT group names" were assigned based on biochemically or genetically validated activities from the Uniprot database query (Dataset S2), though it is important to note that some of them might have been further functionalized in certain taxa. We could not name 15 AT and related groups and hence noted as "uncharacterized (UC)". Overall, class I contained the largest number of 22 groups with 12 AT, 5 non-AT, and 5 UC enzyme groups (*SI Appendix*, Fig. S3) while 6 of 10 groups in class II, 7 of 11 groups in class III, and 2 of 11 groups in class IV as well as 3 groups in sugar AT and alliinases classes were AT groups (*SI Appendix*, Figs. S4–S7). The novel Asp ATs (PF12897, fold-type I) (26, 67), distinct from canonical Asp ATs of class I (PF00155), were found only in two eubacteria and an outgroup of the alliinase class (Fig. 1*C*). The 17 non-AT groups, which belong to AT classes but likely have non-AT activity (e.g., synthases, lyases, and other transferases), were still included in the analysis, as they improved the phylogenetic relationships of ATs (Dataset S4), and some non-AT members can contain ATs, e.g., a Trp AT group found in the alliinases class (68) (*SI Appendix*, Fig. S7). Mapping of AT and related enzymes from four model species—*Escherichia coli*, *Saccharomyces cerevisiae*, *Homo sapiens*, *Arabidopsis thaliana*—suggested complex evolutionary history of AT groups, many of which appear to have been differentially lost in multiple taxa (*SI Appendix*, Figs. S3–S7 and Dataset S4).

To further trace the deep evolutionary history of 62 AT and related groups since LUCA, we mapped and clustered different AT and related enzyme groups based on their presence and absence in the 90 species (Fig. 2 and Dataset S5) and calculated their percent conservation within each taxon (Fig. 2*A*). The most striking finding was that there was no AT group that is absolutely conserved across the ToL, which was unexpected for core metabolic enzymes catalyzing essential reactions. Only one AT group, class IV Ala-glyoxylate ATs (AGTs, *SI Appendix*, Fig. S6)—along with non-AT groups for CoA-utilizing synthases (class I, *SI Appendix*, Fig. S3) and Cys desulfurases (class IV, *SI Appendix*, Fig. S6)—was distributed across almost all taxa though still absent in some species, especially prokaryotes (cluster 1 or CT1 in Fig. 2*B*). Five AT

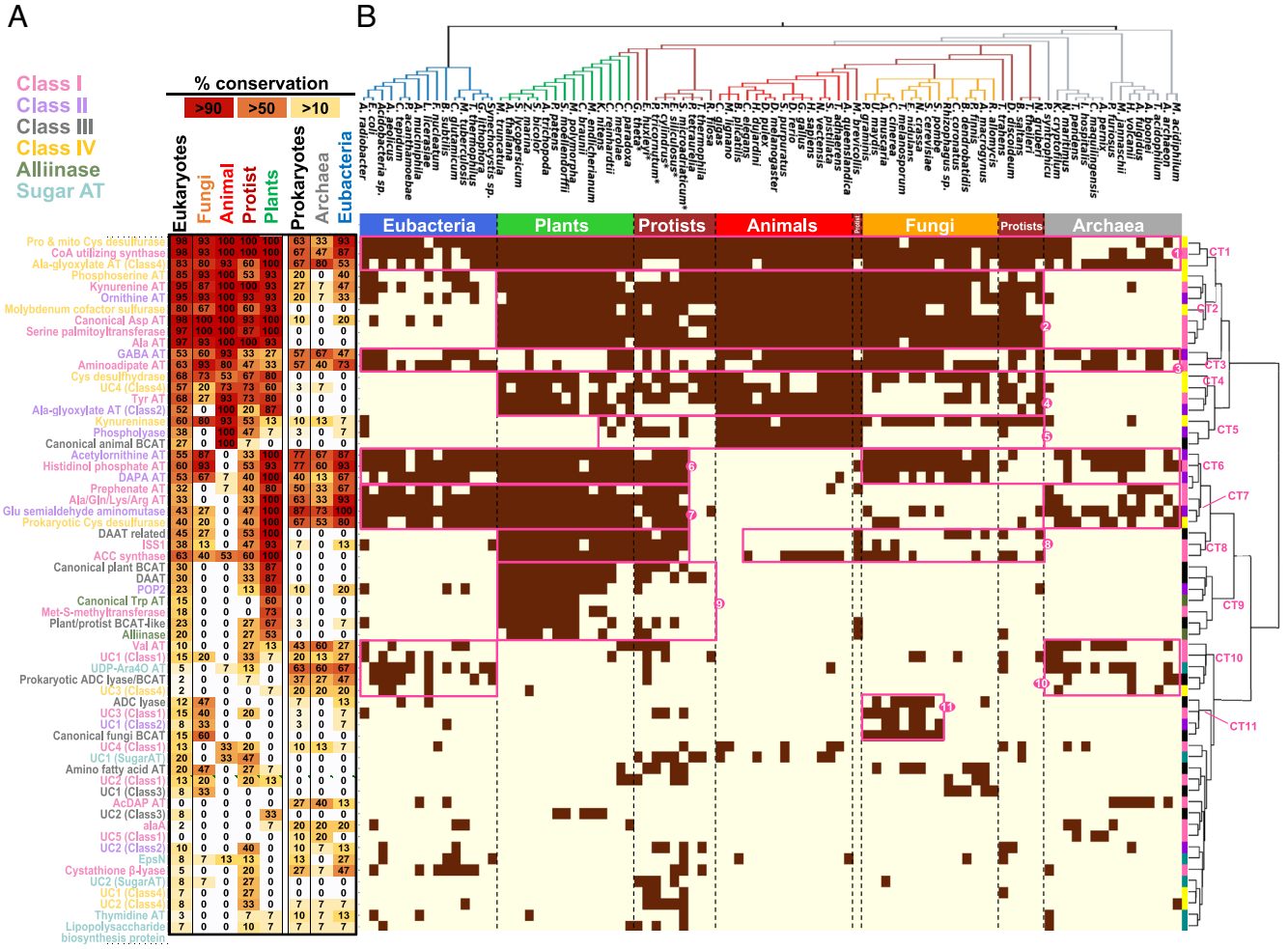

Fig. 2. Poor conservation of AT groups across the ToL due to wide-spread replacement of ATs from distantly related nonorthologous AT groups. (*A*) Percent conservations (with red to yellow background colors) of 62 AT and related groups for different taxonomic groups and ranks. (*B*) AT and related groups were clustered based on the similarity of gene copy numbers within each species. Species were arranged based on the taxonomic relationship at the *Top*: Gray, brown, orange, red, green, and blue depict archaea, protists, fungi, animals, plants, and eubacteria, respectively. Species marked with a star (*) contain secondary plastids. Brown filled boxes indicate that certain enzymes (*Left*) are present in the corresponding species (*Top*). Magenta open boxes highlight 11 clusters (CTs) which are labeled at the corresponding branches of the clustering tree.

groups—Asp ATs, Ala ATs, phosphoserine (PSer) ATs, kynurenine AT, and ornithine ATs (CT2, Fig. 2*B*)—were largely conserved among eukaryotes, but not in prokaryotes. Tyrosine (Tyr) AT and class II AGTs (CT4, Fig. 2*B*) were present in animals and land plants but largely absent in fungi. Although the outcome might vary with the inclusion of other species, further investigations of selected AT groups (e.g., PSer AT, Tyr AT) using the NCBI Landmark model species database, containing high-quality genome assemblies of 27 species across the ToL (including 10 from our 90 species, such as *A. thaliana*, *H. sapiens*, *E. coli*), also yielded consistent results with similar distribution patterns (*SI Appendix*, Fig. S8).

Histidinol-phosphate (HisP) ATs, 7,8-diaminopelargonic acid (DAPA) ATs, and acetylornithine ATs (CT6, Fig. 2*B*) were conserved across the ToL except in animals and nonphotosynthetic protists, owing to the lack of histidine (His), biotin, arginine (Arg) biosynthesis in most of these taxa (69–71). Interestingly, HisP ATs were found in choanoflagellates (e.g., *M. brevicollis*, Fig. 2*B* and Dataset S5), suggesting that de novo His biosynthesis was likely lost after the divergence of animals and choanoflagellates ~600 Mya (72). Ala/glutamine (Gln)/lysine (Lys)/Arg ATs and prephenate ATs, along with glutamate (Glu) semialdehyde aminomutase involved in chlorophyll biosynthesis (73), were highly conserved in bacteria, plants, and photosynthetic protists (CT8, Fig. 2*B*). Some AT groups show highly lineage-specific distribution, which includes animal- and fungi-specific branched-chain amino acid ATs, and land plant-specific Trp ATs that likely functionalized from alliinases (CT5, 11, and 9, respectively, in Fig. 2*B*). Interestingly, ISS1 (or also known as VAS1, class I, *SI Appendix*, Fig. S3) (74) and D-amino acid AT-related enzyme (class II, *SI Appendix*, Fig. S4) groups, whose functions are not fully understood, were highly conserved among all plant species and photosynthetic protists, except diatom *Phaeodactylum tricornutum* (CT9, Fig. 2*B*).

The above AT mapping across the ToL showed that most AT groups are not ubiquitous, unlike other core enzymes carrying out essential metabolic reactions (11, 13, 15, 65, 75–78). Also, closely related ATs (e.g., class II ATs) did not cluster together at all based on taxonomic distribution (Fig. 2). This is likely because AT enzymes from a distant AT group were frequently recruited and replaced existing AT enzymes in certain taxa to carry out essential AT reactions—a phenomenon known as nonorthologous gene displacement (32). One clear example was the enzymes responsible for Ala AT *activity*, which is essential in all organisms (23). In eukaryotes, Ala AT activity is provided mainly by class I Ala ATs (97% conserved among eukaryotes, Fig. 2*A*), and to a lesser extent by the side activities of class II and IV AGTs, and PSer ATs (Fig. 2*C* and *SI Appendix*, Fig. S9*A*). In contrast, in some prokaryotes like *E. coli*, Ala AT activity can be redundantly provided by a number of ATs, e.g., avtA, aspC, tyrB, alaA, serC, and ilvE (79) that are poorly conserved among prokaryotes (Fig. 2 and *SI Appendix*, Fig. S9*A*). Since at least one AT enzyme with Ala AT activity was present for all species analyzed, Ala AT *activity* is conserved across the ToL but mediated by taxon-specific AT *enzymes* (*SI Appendix*, Fig. S9*A*). Additionally, the Tyr AT group (class I) is widespread among eukaryotes but not in many fungi (Fig. 2, *SI Appendix*, Fig. S3, and Dataset S4), whose Tyr AT genes were likely lost and replaced with aminoadipate ATs that belong to a distant clade of class I [e.g., yeast Aro8 and Aro9 (80), *SI Appendix*, Figs. S3 and S9*B*]. Asp AT activity is provided by canonical Asp AT enzymes (class I) in most organisms and additionally by prephenate AT enzymes in many photosynthetic organisms and some prokaryotes (*SI Appendix*, Fig. S9*C*). However, a few eubacteria have noncanonical Asp ATs (PF12897), (26) and we could not identify any known Asp AT enzymes in many archaea species (*SI Appendix*, Fig. S9*C*). Similarly, no clear orthologues were found in land plant genomes for aminoadipate AT enzymes required for Lys metabolism (Fig. 2*B* and *SI Appendix*, Fig. S3). While some of these missing genes could be due to sequencing and/or annotation errors (81, 82), these results overall suggest that there are still unknown ATs responsible for catalyzing these essential AT reactions.

Overall, the global AT mapping across the ToL revealed that essential AT activities were not necessarily carried out by conserved, orthologous enzymes belonging to the same AT class and group in different taxonomic lineages. This likely explains the difficulty in homology-based assignment of AT functions. Instead, we observed many instances of widespread AT enzyme replacement from distantly related, *nonorthologous* AT groups, where ATs from unrelated groups and even different classes were recruited to redundantly or alternatively carry out specific AT reactions. Therefore, in contrast to other core metabolic enzymes, many ATs underwent a complex evolutionary history characterized by the extensive recruitment and replacement of distantly related ATs.

Distantly Related ATs Exhibit Conserved Multisubstrate Specificity. The frequent recruitment and displacement of AT enzymes from distant AT groups may be facilitated by the promiscuity and functional redundancy of AT enzymes among different groups. However, AT substrate specificity has not been extensively studied (23). We therefore experimentally examined the extent of overlap in substrate specificities within the same and among different AT groups. Here, we employed aromatic (Aro) AT activity as a testbed, as it utilizes structurally distinct substrates but is widely detected across distantly related ATs within class I (*SI Appendix*, Table S3), which include the AT groups of ISS1 (83), canonical Asp AT (84), aminoadipate AT (80), HisP AT (85), Tyr AT (86), kynurenine AT (87), and prephenate AT (88) (*SI Appendix*, Fig. S3). As their substrate specificity has not been fully defined, we expressed, purified, and characterized recombinant enzymes from i) canonical Aro AT groups (as positive controls), ii) non-Aro AT groups with no reported canonical Aro ATs, and iii) non-AT group (i.e., *A. thaliana* C-S lyase, *At* SUR1, as a negative control, *SI Appendix*, Table S3). Their substrate specificities were determined by a multisubstrate enzyme assay using 15 keto acceptors with Gln or Glu as an amino donor depending on the enzymes. AT reactions were initially carried out for an extended reaction time with high substrate concentrations. Then, multiple reaction products were analyzed simultaneously using liquid chromatography-mass spectroscopy (LC–MS) (*SI Appendix*, Fig. S10*A*) to calculate their percent conversion of the keto acid substrates to the corresponding amino acid products (Fig. 3*A*).

No Aro AT activity was detected in the Ala AT of *A. thaliana* (*At* AlaAT1) and *E. coli* (*Ec* alaA), nor in *At* SUR1 (Fig. 3*A*), consistent with prior reports (23, 89, 90). In contrast, Aro AT activity was detectable, though at varied degrees, in all of the analyzed non-Aro AT group enzymes (Fig. 3*A* and *SI Appendix*, Table S3), which include *At* HisN6 and kynurenine AT (*At* KAT) that are previously not known to have Aro AT activity in plants (91). This result revealed that Aro AT activity is widespread in many ATs beyond those currently designated as canonical Aro ATs. To our surprise, HisP ATs from *A. thaliana* and *E. coli* (*At* HisN6A, *At* HisN6B, and *Ec* hisC, respectively), having only 32% sequence identity (Fig. 3*A* and *SI Appendix*, Table S4), showed a very similar substrate preference, as reflected by the relative abundance of the produced amino acids: His > Trp > Tyr > phenylalanine (Phe, Fig. 3*A* and *SI Appendix*, Table S5). Similarly, *At* KAT and *Ec* ybdL, having 40% sequence identity, produced the same eight amino acids with a

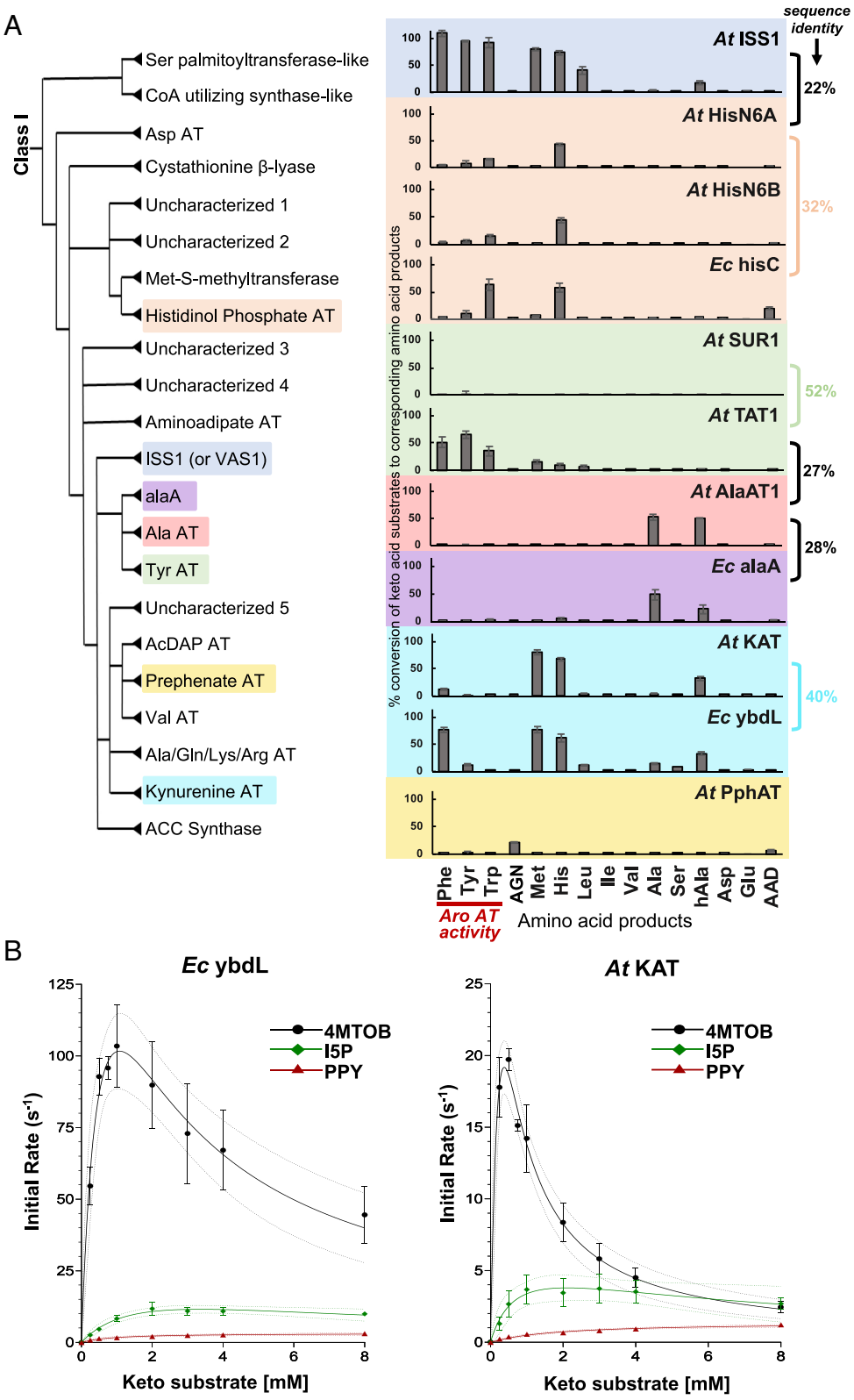

Fig. 3. Conserved multisubstrate specificities across distantly related class I ATs. (*A*) Percent conversion of keto acids to amino acids by Aro ATs and other ATs. Substrate specificities of these ATs, including Aro AT activity (marked by dark red letters and line), were screened in an assay mixture employing a single amino donor (5 mM Gln or Glu) and 15 acceptors (1 mM each). The *X*-axis shows the amino acid products formed by each enzyme. Amino acid standard curves were used to calculate the concentration of the formed amino acids. Molar ratio of amino acid product to the starting keto acid was used to calculate percent conversion, except for Tyr that was calculated based on the decreases in 4-HPP peak area. Each data point is an average of three separate assays (n = 3), except serine which is from a single assay. Error bars show SD among the assays. hAla, homo-Ala; AAD, α-aminoadipate; AGN, arogenate. (*B*) Kinetic characterization of *A. thaliana* (*At*) KAT and *E. coli* (*Ec*) ybdL. Enzymatic activity of *Ec* ybdL and *At* KAT was tested with 10 mM Gln as amino donor and varying concentrations of three prominent keto acid substrates: 4-methylthio-2-oxobutanoic acid (4MTOB), imidazol-5-yl pyruvate (I5P), and phenylpyruvate (PPY). Each data point is an average of three separate assays (*n* = 3). Error bars show SEM. For 4MTOB and I5P, a modified Michaelis–Menten equation that considers substrate inhibition was fitted using nonlinear regression (*SI Appendix*, Fig. S10). For PPY, the standard Michaelis–Menten equation was fitted. Percent amino acid sequence identities between key enzymes are given on the *Right*, and a complete list is given in *SI Appendix*, Table S4.

similar substrate preference, although *Ec* ybdL showed stronger activity with Phe than *At* KAT (Fig. 3 and *SI Appendix*, Table S5). We also detected the conservation of similar substrate promiscuity *across* different groups. All Aro ATs including ones from ISS1, His AT, Tyr AT, and kynurenine AT groups, except prephenate AT, showed activity with Met, His, Leu, and homo-Ala (Fig. 3*A* and *SI Appendix*, Table S5). Therefore, although many ATs show very broad substrate promiscuity, we identified the presence of common signatures of multisubstrate specificity among different ATs even after being separated for billions of years of evolution.

For quantitative comparison, we further conducted kinetic characterization of *Ec* ybdL and *At* KAT in their linear ranges of activity (*SI Appendix*, Fig. S10*B*). Both enzymes had the highest kinetic efficiency with 4-methylthio-2-oxobutanoic acid (4MTOB), followed by imidazol-5-yl pyruvate (I5P) and phenylpyruvate (PPY, keto acids of Met, His, and Phe, respectively, Fig. 3*B* and *SI Appendix*, Fig. S10 *C–H*), which largely agreed with the results of the initial substrate specificity screening (Fig. 3*A*). Interestingly, *Ec* ybdL and *At* KAT showed very similar response curves to these three substrates, except for overall higher turnover rates (k_{cat}) for *Ec* ybdL than *At* KAT (*SI Appendix*, Fig. S10*H*). Additionally, both enzymes showed similar substrate inhibition with 4MTOB and I5P, but not with PPY (Fig. 3*B*), as was also reported for human KAT (92). Importantly, unlike the animal KATs (87) and *Ec* ybdL having kynurenine AT activity, *At* KAT was not capable of using kynurenine as an amino donor (*SI Appendix*, Fig. S11). This result confirmed the lack of *Ec* ybdL contamination in the recombinant *At* KAT preparation and showed functional specialization among kynurenine AT orthologs in different organisms.

Taken together, these results uncovered remarkable similarities of multisubstrate specificity and kinetic properties between ATs from the same groups but taxonomically distant species (e.g., *E. coli* vs. *Arabidopsis*). This conserved promiscuity may also represent their ancestral activity that may date back to LUCA. On the other hand, we detected the overlapping multisubstrate specificity among distantly related, nonorthologous AT groups within class I ATs (Fig. 3*A*), which likely provides functional redundancy and facilitates the AT displacement between distant AT groups (Fig. 2).

The Common Signature of Substrate Specificity among Distantly Related ATs is Mediated Via Distinct Yet Functionally Conserved Residues. To elucidate underlying mechanisms of the overlapping substrate specificity, or functional redundancy, among different AT groups (Fig. 3), we examined whether the presence of certain active site residues or motifs correlates with their substrate specificity. We first obtained experimental or homology-based structures of representative enzymes from each AT group (108 enzymes in total, Dataset S6) and determined the active site residues manually by referring to two crystal structures bound with a hydrophilic Asp [PDB ID: 1ARG (93)] or a hydrophobic Phe [PDB ID: 1W7M (94)] ligand, respectively. The original multiple sequence alignment was then used to calculate the consensus of the structurally conserved (≥10%) active site residues for each AT group (Fig. 4*A*).

As expected (23, 24), the Asp and Lys residues involved in cofactor PLP binding (red arrows in Fig. 4*A*) were highly conserved across all AT groups belonging to PLP-fold type I. The C-terminal Arg residue, which recognizes substrate's carboxylate (yellow arrows in Fig. 4*A*), was mostly conserved with some exceptions, such as POP2 and DAPA AT (class II) that use non-α-amino acid substrates having an increased molecular distance between the transferable amino and carboxyl groups (96, 97). Besides, several other residues were generally well conserved for most PLP-fold type I groups (black arrows in Fig. 4*A*), while each AT class had certain motifs that are conserved among most or only within a subset of AT groups (orange and blue boxes in Fig. 4*A*, respectively). Class IV ATs do not share a ubiquitously conserved motif with other class ATs, consistent with its membership in a different fold-type of PLP enzymes (24, 98) (Fig. 1).

Next, we analyzed whether distantly related AT groups with similar multisubstrate specificity have shared residues or motifs among each other, by comparing the active site residues of class I AT enzymes having Aro AT activity: ISS1, Tyr AT, kynurenine AT groups (Fig. 3), as well as canonical Asp AT and aminoadipate AT groups (80, 83–88). However, the active site sequences of these five AT groups were not more similar to each other than other class I ATs (Fig. 4*A*) and did not share common residues or motifs in their active sites. Therefore, we then employed a structure-based comparison to identify ligand interacting residues by docking Phe-external aldimine complex, as a generic aromatic substrate, to the active sites of their homodimeric complex structures (Fig. 4*B* and *SI Appendix*, Fig. S12). Interestingly, the unique binding pose of Phe in each AT group requires the involvement of several active site residues that were not necessarily conserved at the level of the peptide sequence (Fig. 4*A* and *SI Appendix*, Table S6). For example, nonconserved N-terminal aliphatic amino acids interact with the aromatic ring of Phe in Arabidopsis ISS1 (Met 19), Arabidopsis TAT1 (Ile 13), yeast Aro8 (Leu 25), *E. coli* tyrB (Leu 39), and Arabidopsis HisN6A (Ile 34) (red stars on Fig. 4*B* and *SI Appendix*, Fig. S13). Furthermore, a corresponding residue is absent for human KAT1 and *Streptomyces bingchenggensis* prephenate AT (Sb PphAT), but present for non-Aro ATs such as human and Arabidopsis Ala ATs (Val 64 and Val 82, respectively, *SI Appendix*, Table S6). Similarly, the conserved Tyr near the phosphate group of PLP (*SI Appendix*, Table S6) interacts with both the aromatic ring of Phe and the phosphate group of PLP in Arabidopsis ISS1 (Tyr 64) and Arabidopsis TAT1 (Tyr 71), but only with PLP in the human KAT1 (Tyr 63), yeast Aro8 (Tyr 105), and Arabidopsis HisN6A (Tyr 80) (purple stars on Fig. 4*B*). Instead, the Phe 278, Tyr 251, and Tyr 256 residues interact with the substrate Phe in KAT1, Aro8, and HisN6A, respectively (orange stars on Fig. 4*B*).

We also examined why *Ec* ybdL (99) and *At* KAT had different specificity toward Phe and kynurenine substrates (Fig. 3 and *SI Appendix*, Figs. S10 and S11). Docking of these substrates to these two enzymes showed that both enzymes had nearly identical active site residues (*SI Appendix*, Table S6); however, there were five residues that are distant to the substrate and located around the active site entrance, which appear to make the *Ec* ybdL active site wider with higher solvent accessibility than *At* KAT, likely aiding the access of bulky aromatic substrates into the active site of *Ec* ybdL (*SI Appendix*, Fig. S14 *A* and *B*). Additionally, the N-terminal helix that plays a critical role in the substrate binding of ATs by conformational change (94, 100, 101) had more hydrophobic surface in *At* KAT (I74 and V77) than *Ec* ybdL (A25 and Q28) (*SI Appendix*, Fig. S14 *C* and *D*), potentially restricting the mobility of the N-terminal helix in *At* KAT.

Overall, these sequence and structural analyses of distantly related Aro ATs suggest that they achieve similar multisubstrate specificity using different active residues that are not necessarily conserved at peptide sequences. Additionally, residues that do not directly participate in substrate recognition appear to contribute to AT substrate specificity by influencing overall active site conformation (100).

Discussion

Nitrogen is essential for all life, but its availability significantly varies among different species and in various environments. Therefore, we expect considerable diversity in the number and functionality of AT enzymes that are at the core of the nitrogen metabolic network.

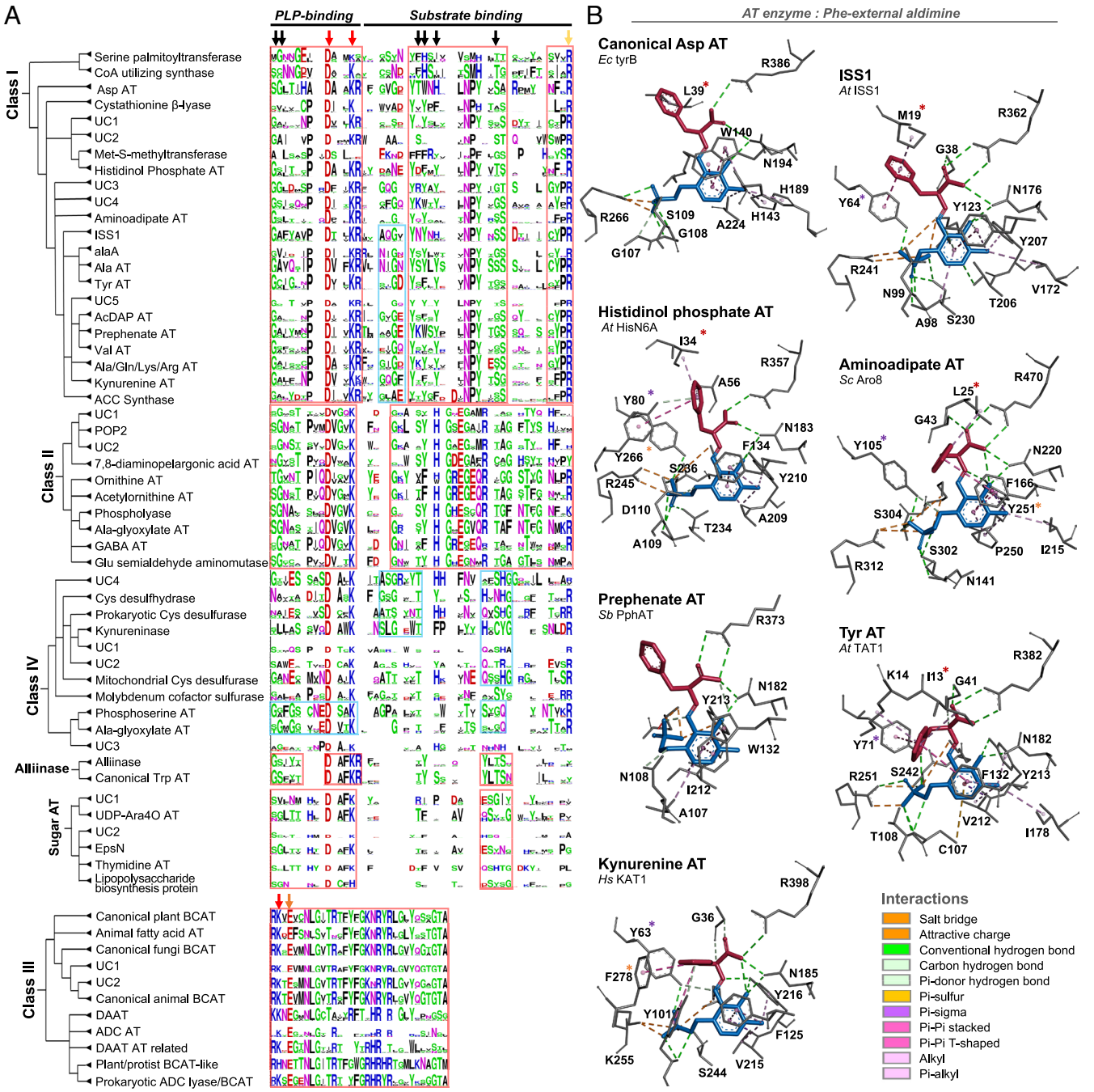

Fig. 4. Distinct but functionally conserved residues underlie the conserved multisubstrate specificity among distantly related ATs. (*A*) Active site residues are extracted from representative enzymes from each AT group and are used to determine the consensus at those residue positions using the original multiple sequence alignment, which is shown in logo style (95). Red and orange arrows show previously known well-conserved residues for PLP-fold type I ATs, and black arrows show additional residues identified in this study. Red and blue boxes show motifs that are well conserved for an entire or a subset of each AT enzyme class, respectively. (*B*) Active site residues that interact with the PLP-Phe aldimine. The portions of the aldimine that correspond to Phe and PLP are shown in red and blue sticks, respectively. Interacting residues are shown in black sticks and labeled with orange letters. The types of interactions are shown with colored dashed lines. Residues having a conserved function in ligand interactions are marked with red, purple, and orange stars, as described in the main text.

However, despite the ~200-fold expansion of proteome sizes from the simplest prokaryotes to land plants and metazoans associated with their metabolic and organismal complexity (65), we found that the overall copy number of AT genes remains largely unchanged. This is in sharp contrast to many of specialized metabolic enzymes that also exhibit substrate promiscuity and underwent tremendous gene expansion to support chemodiversity and ecological adaptation (35–38). Considering the ancient origins of ATs, many new AT copies must have been generated over evolutionary time, as most eukaryotic lineages underwent whole genome duplication (102–106), and horizontal gene transfer events are prevalent among prokaryotes (20, 21). However, the majority of these redundant AT copies were likely lost, resulting in constant AT copy numbers across the ToL. This is likely due to the metabolic cost of maintaining redundant copies involved in core nitrogen metabolism, as compared to relatively low metabolic cost of often conditionally induced specialized metabolism (107). In support of this hypothesis, carbon-sulfur lyases involved in glucosinolate specialized metabolism (90), within one branch of the Tyr AT group (class I), underwent a rapid copy number expansion in *A. thaliana* (*SI Appendix*, Fig. S3) and other Brassicaceae species through gene duplication and neofunctionalization (86, 91). Thus, AT enzymes exhibit characteristics of

both specialized and core metabolic enzymes; they display promiscuity akin to many specialized metabolic enzymes but do not show expand copy numbers similar to other core metabolic enzymes.

Mapping of different AT groups across the ToL showed poor conservation of many AT groups across different organisms (Fig. 2). While certain taxa exhibited the loss of specific AT activity and its associated pathways (such as the absence of HisP AT and His biosynthesis in animals), many AT reactions remain vital across all life forms. Consequently, the absence of orthologous ATs for these crucial activities in certain taxa suggests widespread occurrences of nonorthologous gene displacements (32), wherein distantly related ATs have been recruited to catalyze these indispensable reactions. AT enzymes catalyzing essential AT reactions, such as Ala AT and Tyr AT reactions (*SI Appendix*, Fig. S9), were replaced by an AT from a distantly related, *nonorthologous* group, such as distinct AT groups within the same AT class (i.e., ancient paralogues) or that belong to different AT classes originated from different founder enzymes (i.e., nonhomologues). This is in contrast to other essential core metabolic enzymes that are typically carried out by *orthologous* enzymes (11–17). This is also different from horizontal or endosymbiotic gene transfer of an ortholog from a different taxa, which sometime displaced the existing ortholog and contributed to the "mosaic" evolutionary origins of core metabolic pathways (108, 109). Prior studies documented nonorthologous gene displacements for enzymes involved in some essential biochemical pathways, such as cofactor biosynthesis and nucleotide metabolism, which underscored a major hurdle in reconstructing metabolic pathways based on comparative genomic analyses (32, 110–112). This complex evolutionary history of the AT enzyme family, therefore, highlights the challenges associated with homology-based prediction of AT enzyme functions, calling for the need of additional experimental testing to elucidate and validate AT functions. These findings also suggest that the recruitment of ATs from nonorthologous groups likely led to diverse architecture and functionality of the nitrogen metabolic networks that exist in different extant organisms.

What are the underlying mechanisms that potentially contributed to the complex history of AT evolution? The detailed analyses of distantly related class I Aro ATs (Fig. 3*A* and *SI Appendix*, Tables S3–S5) identified that their multisubstrate specificity is highly conserved among ATs from taxonomically distant species, such as kynurenine ATs from *A. thaliana* (plants) and *E. coli* (eubacteria); they also showed striking similarities for their substrate inhibition (Fig. 3*B* and *SI Appendix*, Figs. S10 and S11) (92). Additionally, a common signature of multisubstrate specificity with His, Met, and Leu, besides aromatic amino acids, was conserved among distantly related Aro ATs that belong to different AT groups (Fig. 3*A*). These overlapping substrate specificities provide functional redundancy and hence mutation tolerance, allowing a certain AT to be lost or be neofunctionalized without gaining a new gene copy—which would be detrimental to other essential core metabolic enzymes. This can be seen in *E. coli* quadruple mutant that lacks all three major Ala ATs (*alaA, avtA, alaC*) plus *serC* having Ala AT activity (*SI Appendix*, Fig. S9) but is still not Ala auxotrophic (89). Therefore, the extreme functional redundancy, provided by the overlapping multisubstrate specificities across distantly related ATs, likely enabled the widespread AT enzyme replacement without expanding AT gene copy numbers.

Prior studies propose that AT promiscuity likely arose from mechanistic constraints of transamination reactions that require sequential binding and transformation of two substrates, an amino donor and acceptor (Fig. 1*A* and *SI Appendix*, Fig. S1*A*) (23, 113). These often distinct molecules need to be accommodated in the AT active site, such as through the large-scale rearrangement of the hydrogen bond network or at hydrophilic and hydrophobic pockets (23, 113). Homology-based comparisons of Aro AT active sites failed to identify conserved motifs or residues (Fig. 4*A* and *SI Appendix*, Table S6). Instead, we found functional versatility of AT active sites in recruiting new or already existing residues for novel functions (Fig. 4*B* and *SI Appendix*, Fig. S13), which can act as a reservoir for the emergence of new AT functions. For example, *Mycobacterium tuberculosis* has an Aro AT (known as Ar AT) within the HisP AT group (85) (*SI Appendix*, Fig. S3), which might have arisen from the positive selection of weak Aro AT activity of an ancestral HisP AT to become a bona fide Aro AT enzyme. Additionally, our analysis suggests that residues outside of the active site can induce conformational changes to alter the topography of the active site (*SI Appendix*, Fig. S14), which might have also contributed to the emerging specificity of Aro ATs. While more AT enzymes need to be characterized from other AT groups and classes, we hypothesize that the overlapping substrate specificity of distantly related ATs can convergently evolve in distantly related AT groups through recruitment of distinct but functionally conserved residues.

In summary, the multifunctionality of ATs and their functional redundancy increase the mutation tolerance of AT genes and allow replacement of distinct ATs that belong to different AT groups or even classes, without necessarily expanding AT gene copy numbers per species. The mixing and matching of AT enzymes from different AT groups and classes, having different side activities and properties, likely contributed to the robust and diverse nitrogen metabolic networks present throughout the ToL (28, 31). This ToL-scale map of the entire AT family, coupled with the use of recently developed mass spectrometry (MS)-based high-throughput AT assays (64, 114), now enables systematic determination of the multisubstrate specificity of various AT enzymes. This will allow further mechanistic studies of the AT functions and multisubstrate specificity and enable AI-based prediction of AT sequence–structure–function relationships. Well-defined AT functionalities from different organisms will allow us to define diverse nitrogen metabolic networks that likely operate in various extant organisms dealing with different nitrogen availabilities and demands. The fundamental understanding of AT functions and nitrogen metabolism will in turn provide rational strategies to develop therapeutics against pathogens and metabolic disorders, redesign nitrogen metabolic networks through synthetic biology, and enhance nitrogen use efficiency for sustainable crop production.

Materials and Methods

Selection of Species. Species used in this study were selected based on the availability of high-quality proteomes (designated as "reference proteome" by Uniport, except some plant species queried from Phytozome). Taxon sampling was performed for each kingdom and protists to include species that would best represent the diversity of the investigated taxonomic group. It is important to note that the selection of the number of species from each taxon is not impartial, as it overrepresents eukaryotic lineages, particularly archaeplastida, holozoa, and holomycota. However, since species from these groups are also disproportionately represented in scientific research (115), we concluded that their overrepresentation would benefit a broader audience.

Acquisition of AT Sequences. To identify putative AT sequences, we performed Pfam analysis. All protein sequences were downloaded in August 2021 for 90 species from three public databases: Phytozome (42) and Phycocosm (44) for plants and UniProt (116) for the others. To avoid redundant protein sequences, we used primary transcripts and nonredundant proteome data for plants and the others, respectively (Dataset S7). The proteins shorter than 250 amino acids were excluded owing to the potential for nonfunctional genes. The sequences after length

filtration proceed to Pfam annotation by HMMER v3.3.1 (hmmscan with a default setting; E-value cutoff = 0.01) (117) using Pfam profile hidden Markov models obtained from Pfam v35.0 (46). In the case of multiple hits for the same region in the protein sequence, we chose the Pfam with a lower E-value. Then, we obtained 2,938 putative AT sequences (Dataset S2) by searching Pfam IDs in which proteins display AT activity: PF00155, PF00202, PF00266, PF04864, PF01041, PF12897 (PLP fold type I), and PF01063 (PLP fold type IV). For the extremely long proteins (>1,200 amino acids), we extracted the only AT Pfam domain region. In the case that multiple AT Pfam domains exist, the longer AT region was selected.

Multiple Sequence Alignment of AT Sequences and the Phylogenetic Tree Construction. For structure-guided multiple sequence alignment, amino acid sequences were aligned by a MAFFT-DASH (49) using a BLOSUM62 scoring matrix (gap opening penalty = 1.53). We assumed closer associations among the proteins belonging to the same Pfam. Thus, all putative AT sequences were divided into seven groups according to Pfam IDs, followed by multiple sequence alignment for each group with iterative refinement methods (L-INS-i for PF01041, PF04864, and PF12897; FFT-INS-i for PF00155, PF00202, PF0266, and PF01063). Then, considering the evolutionary divergence of PLP fold type I, we merged the alignments for six Pfam (i.e., PF00155, PF00202, PF00266, PF04864, PF01041, and PF12897) into a single multiple sequence alignment for PLP fold type I with a progressive method (FFT-NS-2). The alignment for PF01063 was regarded as the one for PLP fold type IV. For further analysis, we deleted DASH sequences from the alignments to avoid misinterpretation.

To construct phylogenetic trees for large alignments, we employed an approximately maximum likelihood method. The alignments for PLP fold types I and IV were analyzed by FastTree v2.1.11 (51) with LG+CAT model and the option "-mlacc 2 -slownni -spr 4 -pseudo". To avoid a potential issue of a local optima described in the previous study (66), we generated 40 FastTree trees using 20 random and 20 parsimony starting trees prepared by RAxML-NG (v1.2.0) with a JTT+G model, and then selected the most likelihood trees. For the PLP-fold type I, we additionally built 40 FastTree trees for each class (e.g., class I ATs, Alliinases, Sugar ATs) and then selected the most likelihood trees. To evaluate the reliability of each split in the tree, we used local support values (118) computed in FastTree from 1,000 resamples. We analyzed the tree and modified the tree topology using an ETE 3 Toolkit (119) in a Python environment (v3.6.13). To identify the highly supported phylogenetic branches, the nodes with low supporting values (<90%) were deleted and then integrated into parental nodes. The condensed phylogenetic trees were manually drawn using Microsoft PowerPoint with local support values as branch support.

The functions of AT and related groups were inferred from the publicly available experimental data. First, we picked four species whose enzymes have been well defined (i.e., *H. sapiens*, *A. thaliana*, *S. cerevisiae*, and *E. coli*), and then inquired about the functions of AT candidates from literature search. Also, we collected additional 68 enzymes whose experimental data are available in the UniProtKB (116), the BRENDA (120), or the SABIO-RK (121) databases. This information was utilized to annotate names for the AT and related groups. We labeled the group with UC when any group member was unavailable in the collection. Only enzyme clades that have more than five sequences from at least four species or three kingdoms with a bootstrap value higher than 90% were assigned AT groups names. Sequences that did not belong to any groups were indicated as "unassigned".

Mapping of AT Groups on the Taxonomic Tree. We computed how many AT-coding genes each species carries for 62 AT and related groups. This dataset was used for two different analyses to examine taxonomic conservation and distribution. For the taxonomic conservation, the sum of the gene counts was calculated in each kingdom for each AT and related group. The values were used to generate a heatmap using Microsoft Excel.

To investigate the taxonomic distribution, we first transformed the gene counts into the possession (i.e., 0: absence, >1: presence) of individual AT and related groups. A heatmap was generated with columns being clustered (distance metric: "euclidean", linkage method: "ward") by the seaborn package v0.11.2 (122) in a Python environment, whereas rows were arranged phylogenetically by referring to the previous studies (39–41, 43–45) to be able to trace the evolutionary history of individual ATs.

The taxonomic distribution of selected AT groups was separately examined using the NCBI Landmark model species database, as described in *SI Appendix, SI Method*.

Analysis of Subcellular Localization of AT and Related Enzymes. All AT candidates determined by Pfam analysis were analyzed for their predicted subcellular localization. The presence of N-terminal presequences for subcellular localization was analyzed using TargetP-2.0 server (123). We focused on the mitochondrial and chloroplast transit peptide and ignored the presence of signal peptides for secretory pathways. Based on this prediction, the AT candidates were classified into 3 groups according to subcellular location: mitochondria, chloroplast, and others.

AT Enzyme Assays. AT enzymes were expressed and purified as described previously (114) and in *SI Appendix, SI Method*. Multisubstrate specificity screenings as well as the enzyme kinetic analyses of ATs were conducted as described previously (114) and in *SI Appendix, SI Method*.

Determination of the Variance of Active Site Residues. We initially picked 111 model proteins over 62 AT-like groups (1 to 4 enzymes from each group) for structural analysis. The crystal structures of 57 proteins were obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (124). For the other proteins, homodimeric structure models were generated by AI-based structure prediction using AlphaFold v2.1.0 (125, 126). The predicted structures were manually checked whether the dimeric interfaces and the active site formed properly. In the case of a distorted structure or collapsed active site, we constructed another structure model by conventional homology modeling using SWISS-MODEL (127). When the models were still poor, they were excluded from further analyses. Finally, we obtained the structures for 107 proteins over 56 AT groups (Dataset S6). Then, the structures were aligned with each other using the SALIGN module of MODELLER v10.2 (128) and then used for active site analysis. Active site residues were determined manually referring to the protein–ligand complex structures determined previously with Asp and Phe (PDB ID: 1ARG and 1W7M, respectively) (93, 94). The corresponding positions of the active site residues were examined in multiple sequence alignments and minor residues (i.e., <10% conservation among model proteins) were removed. As a result, we obtained 50 and 32 residues for PLP fold types I and IV, respectively. Next, the putative active site residues were extracted from multiple sequence alignments for 2,938 ATs. The lists of active site residues were divided according to AT-like groups, followed by generation of consensus sequences using WebLogo (95).

Molecular Docking Studies. The docking simulations of external aldimine intermediates were performed against 9 ATs (Dataset S8) using AutoDock v4.2.6 (129) as described in *SI Appendix, SI Method*.

Data, Materials, and Software Availability. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank Drs. Prashant Sharma, Garret Suen and Karthik Anantharaman for helping the selection of animal, eubacteria, archaea species, respectively, while Drs. Anne Pringle and Dr. Yen Wen Wang for the selection of fungi species to be used in the ToL analysis. This work was supported by the US Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0020390 and DE-AC02-05CH11231), the Joint Genome Institute award no. CSP-503757, as well as the U.S. NSF award PGRP-IOS-2312181. The work (10.46936/10.25585/60001160) conducted by the DOE Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of DOE operated under Contract No. DE-AC02-05CH11231.

Author affiliations: [a]Department of Botany, University of Wisconsin-Madison, Madison, WI 53706; [b]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Department of Biotechnology, Konkuk University, Chungju 27478, South Korea; [d]The US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [e]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [f]Center for Advanced Bioenergy and Bioproducts Innovation, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [g]Global Center for Food, Land, and Water Resources, Research Faculty of Agriculture, Hokkaido University, Hokkaido, Japan 060-8589; and [h]Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Tokyo 183-8538, Japan

1. M. Yčas, On earlier states of the biochemical system. *J. Theoret. Biol.* **44**, 145–160 (1974).
2. R. A. Jensen, Enzyme recruitment in evolution of new function. *Ann. Rev. Microbiol.* **30**, 409–425 (1976).
3. M. P. Ferla, J. L. Brewster, K. R. Hall, G. B. Evans, W. M. Patrick, Primordial-like enzymes from bacteria with reduced genomes. *Mol. Microbiol.* **105**, 508–524 (2017).
4. A. Wagner, Energy costs constrain the evolution of gene expression. *J. Exp. Zool. Part B Mol. Dev. Evol.* **308B**, 322–324 (2007).
5. H. Nam *et al.*, Network context and selection in the evolution to enzyme specificity. *Science* **337**, 1101–1104 (2012).
6. N. Noffke, D. Christian, D. Wacey, R. M. Hazen, Microbially induced sedimentary structures recording an ancient ecosystem in the ca. 3.48 billion-year-old dresser formation, Pilbara, Western Australia. *Astrobiology* **13**, 1103–1124 (2013).
7. E. A. Bell, P. Boehnke, T. M. Harrison, W. L. Mao, Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14518–14521 (2015).
8. H. C. Betts *et al.*, Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).
9. M. A. Ditzler, M. Popović, T. Zajkowski, "Chapter 5–From building blocks to cells" in *New Frontiers in Astrobiology*, R. Thombre, P. Vaishampayan, Eds. (Elsevier, 2022), pp. 111–133, 10.1016/B978-0-12-824162-2.00010-5.
10. A. P. Pandurangan, J. Stahlhacke, M. E. Oates, B. Smithers, J. Gough, The SUPERFAMILY 2.0 database: A significant proteome update and a new webserver. *Nucleic Acids Res.* **47**, D490–D494 (2019).
11. A. S. Petrov *et al.*, Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10251–10256 (2014).
12. G. M. Nagel, R. F. Doolittle, Phylogenetic analysis of the aminoacyl-tRNA synthetases. *J. Mol. Evol.* **40**, 487–498 (1995).
13. G. P. Fournier, E. J. Alm, Ancestral reconstruction of a Pre-LUCA aminoacyl-tRNA synthetase ancestor supports the late addition of Trp to the genetic code. *J. Mol. Evol.* **80**, 171–185 (2015).
14. M. A. Rubio Gomez, M. Ibba, Aminoacyl-tRNA synthetases. *RNA* **26**, 910–936 (2020).
15. K. S. Smith, C. Jakubzick, T. S. Whittam, J. G. Ferry, Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15184–15189 (1999).
16. N. D. Rawlings, A. J. Barrett, Evolutionary families of peptidases. *Biochem. J.* **290**, 205–218 (1993).
17. N. D. Rawlings, A. J. Barrett, A. Bateman, MEROPS: The peptidase database. *Nucleic Acids Res.* **38**, D227–233 (2010).
18. M. C. Weiss, M. Preiner, J. C. Xavier, V. Zimorski, W. F. Martin, The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* **14**, e1007518 (2018).
19. S. D. Copley, Evolution of new enzymes by gene duplication and divergence. *FEBS J.* **287**, 1262–1283 (2020).
20. H. Ochman, J. G. Lawrence, E. A. Groisman, Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
21. F. D. K. Tria, W. F. Martin, Gene duplications are at least 50 times less frequent than gene transfers in prokaryotic genomes. *Genome Biol. Evol.* **13**, evab224 (2021).
22. H. Maeda, H. Yoo, N. Dudareva, Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nat. Chem. Biol.* **7**, 19–21 (2011).
23. K. Koper, S.-W. Han, D. C. Pastor, Y. Yoshikuni, H. A. Maeda, Evolutionary origin and functional diversification of aminotransferases. *J. Biol. Chem.* **298**, 102122 (2022), 10.1016/j.jbc.2022.102122.
24. P. K. Mehta, T. I. Hale, P. Christen, Aminotransferases: Demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* **214**, 549–561 (1993).
25. K. Qian *et al.*, Hepatic ALT isoenzymes are elevated in gluconeogenic conditions including diabetes and suppressed by insulin at the protein level. *Diabetes Metab. Res. Rev.* **31**, 562–571 (2015).
26. R. S. Jansen *et al.*, Aspartate aminotransferase Rv3722c governs aspartate-dependent nitrogen metabolism in Mycobacterium tuberculosis. *Nat. Commun.* **11**, 1960 (2020).
27. Z.-N. Ling *et al.*, Amino acid metabolism in health and disease. *Signal Transduct. Target Ther.* **8**, 345 (2023).
28. H. Schulz-Mirbach *et al.*, On the flexibility of the cellular amination network in E coli. *Elife* **11**, e77492 (2022).
29. A. K. Shrawat, R. T. Carroll, M. DePauw, G. J. Taylor, A. G. Good, Genetic engineering of improved nitrogen use efficiency in rice by the tissue-specific expression of alanine aminotransferase. *Plant Biotechnol. J.* **6**, 722–732 (2008).
30. M. D. Toney, Aspartate aminotransferase: An old dog teaches new tricks. *Arch. Biochem. Biophys.* **544**, 119–127 (2014).
31. A. Agapova *et al.*, Flexible nitrogen utilisation by the metabolic generalist pathogen Mycobacterium tuberculosis. *Elife* **8**, e41129 (2019).
32. E. V. Koonin, A. R. Mushegian, P. Bork, Non-orthologous gene displacement. *Trends Genet.* **12**, 334–336 (1996).
33. G. D. Moghe, R. L. Last, Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism1. *Plant Physiol.* **169**, 1512–1523 (2015).
34. H. Kusano *et al.*, Evolutionary developments in plant specialized metabolism, exemplified by two transferase families. *Front. Plant Sci.* **10**, 794 (2019).
35. J.-K. Weng, R. N. Philippe, J. P. Noel, The rise of chemodiversity in plants. *Science* **336**, 1667–1670 (2012).
36. M. Mizutani, D. Ohta, Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* **61**, 291–315 (2010).
37. Q. Jia *et al.*, Origin and early evolution of the plant terpene synthase family. *Proc. Natl Acad. Sci. U.S.A.* **119**, e2100361119 (2022).
38. J. P. Torres, E. W. Schmidt, The biosynthetic diversity of the animal world. *J. Biol. Chem.* **294**, 17684–17692 (2019).
39. Q. Zhu *et al.*, Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **10**, 5477 (2019).
40. T. Cavalier-Smith, E.E.-Y. Chao, Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). *Protoplasma* **257**, 621–753 (2020).
41. N. Pires, L. Dolan, Early evolution of bHLH proteins in plants. *Plant Signal. Behav.* **5**, 911–912 (2010).
42. D. M. Goodstein *et al.*, Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
43. C. W. Dunn, G. Giribet, G. D. Edgecombe, A. Hejnol, Animal phylogeny and its evolutionary implications. *Ann. Rev. Ecol. Evol. Syst.* **45**, 371–395 (2014).
44. I. V. Grigoriev *et al.*, MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**, D699–D704 (2014).
45. F. Burki, A. J. Roger, M. W. Brown, A. G. B. Simpson, The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
46. J. Mistry *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
47. K. Illergård, D. H. Ardell, A. Elofsson, Structure is three to ten times more conserved than sequence–A study of structural response in protein cores. *Proteins* **77**, 499–508 (2009).
48. L. Holm, C. Sander, Mapping the protein universe. *Science* **273**, 595–603 (1996).
49. J. Rozewicki, S. Li, K. M. Amada, D. M. Standley, K. Katoh, MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res.* **47**, W5–W10 (2019).
50. R. D. Finn *et al.*, Pfam: Clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
51. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2–Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
52. S. D. Breazeale, A. A. Ribeiro, C. R. H. Raetz, Origin of lipid a species modified with 4-amino-4-deoxy-l-arabinose in polymyxin-resistant mutants of Escherichia coli: An aminotransferase (ArnB) that generates udp-4-amino-4-deoxy-l-arabinose*. *J. Biol. Chem.* **278**, 24731–24739 (2003).
53. B. W. Noland *et al.*, Structural studies of Salmonella typhimurium ArnB (PmrH) aminotransferase: A 4-amino-4-deoxy-L-arabinose lipopolysaccharide-modifying enzyme. *Structure* **10**, 1569–1580 (2002).
54. R. Amann *et al.*, Obligate intracellular bacterial parasites of acanthamoebae related to Chlamydia spp. *Appl. Environ. Microbiol.* **63**, 115–121 (1997).
55. B. J. Baker *et al.*, Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8806–8811 (2010).
56. T. Heimerl *et al.*, A complex endomembrane system in the archaeon Ignicoccus hospitalis tapped by Nanoarchaeum equitans. *Front. Microbiol.* **8**, 1072 (2017).
57. X. Huang *et al.*, Ancestral Genomes: A resource for reconstructed ancestral genes and genomes across the tree of life. *Nucleic Acids Res.* **47**, D271–D279 (2019).
58. A. H. Liepman, L. J. Olsen, Peroxisomal alanine: Glyoxylate aminotransferase (AGT1) is a photorespiratory enzyme with multiple substrates in *Arabidopsis thaliana*. *Plant J.* **25**, 487–498 (2001).
59. D. Von Wettstein, S. Gough, C. Kannangara, Chlorophyll biosynthesis. *Plant Cell* **7**, 1039–1057 (1995).
60. M. Lu *et al.*, Role of the malate–aspartate shuttle on the metabolic response to myocardial ischemia. *J. Theoret. Biol.* **254**, 466–475 (2008).
61. Q. Han, H. Robinson, T. Cai, D. A. Tagle, J. Li, Biochemical and structural characterization of mouse mitochondrial aspartate aminotransferase, a newly identified kynurenine aminotransferase-IV. *Biosci. Rep.* **31**, 323–332 (2011).
62. R.-Z. Yang, G. Blaileanu, B. C. Hansen, A. R. Shuldiner, D.-W. Gong, cDNA cloning, genomic structure, chromosomal mapping, and functional expression of a novel human alanine aminotransferase. *Genomics* **79**, 445–450 (2002).
63. A. N. Stepanova *et al.*, TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* **133**, 177–191 (2008).
64. M. de Raad *et al.*, Mass spectrometry imaging-based assays for aminotransferase activity reveal a broad substrate spectrum for a previously uncharacterized enzyme. *J. Biol. Chem.* **299**, 102939 (2023), 10.1016/j.jbc.2023.102939.
65. M. E. Rebeaud, S. Mallik, P. Goloubinoff, D. S. Tawfik, On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020885118 (2021).
66. X. Zhou, X.-X. Shen, C. T. Hittinger, A. Rokas, Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Mol. Biol. Evol.* **35**, 486–503 (2018).
67. H. F. Son, K.-J. Kim, Structural insights into a novel class of aspartate aminotransferase from Corynebacterium glutamicum. *PLoS One* **11**, e0158402 (2016).
68. V. P. Carrillo-Carrasco, J. Hernandez-Garcia, S. K. Mutte, D. Weijers, The birth of a giant: Evolutionary insights into the origin of auxin responses in plants. *EMBO J.* **42**, e113018 (2023).
69. G. Wu, Amino acids: Metabolism, functions, and nutrition. *Amino Acids* **37**, 1–17 (2009).
70. G. R. F. Davis, Essential dietary amino acids for growth of larvae of the yellow mealworm, Tenebrio molitor L. *J. Nutr.* **105**, 1071–1075 (1975).
71. L. M. Fitzgerald, A. M. Szmant, Biosynthesis of 'essential' amino acids by scleractinian corals. *Biochem. J.* **322**, 213–221 (1997).
72. N. King *et al.*, The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451**, 783–788 (2008).
73. L. L. Ilag, A. M. Kumar, D. Söll, Light regulation of chlorophyll biosynthesis at the level of 5-aminolevulinate formation in Arabidopsis. *Plant Cell* **6**, 265–275 (1994).
74. Z. Zheng *et al.*, Coordination of auxin and ethylene biosynthesis by the aminotransferase VAS1. *Nat. Chem. Biol.* **9**, 244–246 (2013).
75. P. M. Shih *et al.*, Biochemical characterization of predicted Precambrian RuBisCO. *Nat. Commun.* **7**, 10382 (2016).
76. C. Hsiao, S. Mohan, B. K. Kalahar, L. D. Williams, Peeling the onion: Ribosomes are ancient molecular fossils. *Mol. Biol. Evol.* **26**, 2415–2425 (2009).
77. V. Alva, M. Ammelburg, J. Söding, A. N. Lupas, On the origin of the histone fold. *BMC Struct. Biol.* **7**, 17 (2007).
78. H. Cerutti, J. A. Casas-Mollano, On the origin and functions of RNA-mediated silencing: From protists to man. *Curr. Genet.* **50**, 81–99 (2006).
79. T. Iwasaki *et al.*, Escherichia coli amino acid auxotrophic expression host strains for investigating protein structure–function relationships. *J. Biochem.* **169**, 387–394 (2021).
80. A. Urrestarazu, S. Vissers, I. Iraqui, M. Grenson, Phenylalanine- and tyrosine-auxotrophic mutants of Saccharomyces cerevisiae impaired in transamination. *Mol. Gen. Genet.* **257**, 230–237 (1998).
81. E. Martínez-Carranza *et al.*, Variability of bacterial essential genes among closely related bacteria: The case of Escherichia coli. *Front. Microbiol.* **9**, 1059 (2018).

82. R. L. Charlebois, W. F. Doolittle, Computing prokaryotic gene ubiquity: Rescuing the core from extinction. *Genome Res.* **14**, 2469–2477 (2004).
83. M. Pieck *et al.*, Auxin and tryptophan homeostasis are facilitated by the ISS1/VAS1 aromatic aminotransferase in Arabidopsis. *Genetics* **201**, 185–199 (2015).
84. C. Mavrides, W. Orr, Multispecific aspartate and aromatic amino acid aminotransferases in Escherichia coli. *J. Biol. Chem.* **250**, 4128–4133 (1975).
85. N. Nasir, A. Anant, R. Vyas, B. K. Biswal, Crystal structures of Mycobacterium tuberculosis HspAT and ArAT reveal structural basis of their distinct substrate specificities. *Sci. Rep.* **6**, 18880 (2016).
86. M. Wang, K. Toda, H. A. Maeda, Biochemical properties and subcellular localization of tyrosine aminotransferases in Arabidopsis thaliana. *Phytochemistry* **132**, 16–25 (2016).
87. Q. Han, T. Cai, D. A. Tagle, J. Li, Structure, expression, and function of kynurenine aminotransferases in human and rodent brains. *Cell Mol. Life Sci.* **67**, 353–368 (2010).
88. C. Dornfeld *et al.*, Phylobiochemical characterization of class-Ib aspartate/prephenate aminotransferases reveals evolution of the plant arogenate phenylalanine pathway. *The Plant Cell* **26**, 3101–3114 (2014).
89. S. H. Kim, B. L. Schneider, L. Reitzer, Genetics and regulation of the major enzymes of alanine synthesis in Escherichia coli. *J. Bacteriol.* **192**, 5304–5311 (2010).
90. M. D. Mikkelsen, P. Naur, B. A. Halkier, Arabidopsis mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J.* **37**, 770–777 (2004).
91. M. Wang, H. A. Maeda, Aromatic amino acid aminotransferases in plants. *Phytochem. Rev.* **17**, 131–159 (2018).
92. Q. Han, J. Li, J. Li, pH dependence, substrate specificity and inhibition of human kynurenine aminotransferase I. *Eur. J. Biochem.* **271**, 4804–4814 (2004).
93. R. Graber *et al.*, Changing the reaction specificity of a pyridoxal-5′-phosphate-dependent enzyme. *Eur. J. Biochem.* **232**, 686–690 (1995).
94. F. Rossi, Q. Han, J. Li, J. Li, M. Rizzi, Crystal structure of human kynurenine aminotransferase I. *J. Biol. Chem.* **279**, 50214–50220 (2004).
95. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
96. R. Palanivelu, L. Brass, A. F. Edlund, D. Preuss, Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels. *Cell* **114**, 47–59 (2003).
97. G. L. Stoner, M. A. Eisenberg, Purification and properties of 7, 8-diaminopelargonic acid aminotransferase. *J. Biol. Chem.* **250**, 4029–4036 (1975).
98. P. Christen, P. K. Mehta, From cofactor to enzymes. The molecular evolution of pyridoxal-5′-phosphate-dependent enzymes. *The Chem. Rec.* **1**, 436–447 (2001).
99. M. Dolzan *et al.*, Crystal structure and reactivity of YbdL from Escherichia coli identify a methionine aminotransferase function. *FEBS Lett.* **571**, 141–146 (2004).
100. Q. Han, H. Robinson, T. Cai, D. A. Tagle, J. Li, Structural insight into the inhibition of human kynurenine aminotransferase I/glutamine transaminase K. *J. Med. Chem.* **52**, 2786–2793 (2009).
101. Y. Fukumoto *et al.*, Structural and functional role of the amino-terminal region of porcine cytosolic aspartate aminotransferase: Catalytic and structural properties of enzyme derivatives truncated on the amino-terminal side. *J. Biol. Chem.* **266**, 4187–4193 (1991).
102. J. W. Clark, P. C. J. Donoghue, Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* **23**, 933–945 (2018).
103. W. Albertin, P. Marullo, Polyploidy in fungi: Evolution after whole-genome duplication. *Proc. R Soc. B Biol. Sci.* **279**, 2497–2509 (2012).
104. J. B. Ahrens, J. Nunez-Castilla, J. Siltberg-Liberles, Evolution of intrinsic disorder in eukaryotic proteins. *Cell. Mol. Life Sci.* **74**, 3163–3174 (2017).
105. L. Z. Holland, D. Ocampo Daza, A new look at an old question: When did the second whole genome duplication occur in vertebrate evolution? *Genome Biol.* **19**, 209 (2018).
106. Z. Li *et al.*, Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4713–4718 (2018).
107. A. Dadras *et al.*, Accessible versatility underpins the deep evolution of plant specialized metabolism. *Phytochem Rev.* (2023), 10.1007/s11101-023-09863-2.
108. H. A. Maeda, A. R. Fernie, Evolutionary history of plant metabolism. *Annu. Rev. Plant Biol.* **72**, 185–216 (2021).
109. E. C. M. Nowack *et al.*, Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of Paulinella chromatophora. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12214–12219 (2016).
110. V. de Crécy-Lagard, Variations in metabolic pathways create challenges for automated metabolic reconstructions: Examples from the tetrahydrofolate synthesis pathway. *Comput. Struct. Biotechnol. J.* **10**, 41–50 (2014).
111. R. Denise, J. Babor, J. A. Gerlt, V. de Crécy-Lagard, Pyridoxal 5′-phosphate synthesis and salvage in Bacteria and Archaea: Predicting pathway variant distributions and holes. *Microb. Genom.* **9**, mgen000926 (2023).
112. E. Dembech *et al.*, Identification of hidden associations among eukaryotic genes through statistical analysis of coevolutionary transitions. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2218329120 (2023).
113. K. Hirotsu, M. Goto, A. Okamoto, I. Miyahara, Dual substrate recognition of aminotransferases. *Chem. Rec.* **5**, 160–172 (2005).
114. K. Koper, S. Hataya, A. G. Hall, T. E. Takasuka, H. A. Maeda, Biochemical characterization of plant aromatic aminotransferases. *Methods Enzymol.* **680**, 35–83 (2023).
115. M. R. Dietrich, R. A. Ankeny, P. M. Chen, Publication trends in model organism research. *Genetics* **198**, 787–794 (2014).
116. The UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
117. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
118. H. Shimodaira, M. Hasegawa, Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114 (1999).
119. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
120. A. Chang *et al.*, BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res.* **49**, D498–D508 (2021).
121. U. Wittig, M. Rey, A. Weidemann, R. Kania, W. Müller, SABIO-RK: An updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Res.* **46**, D656–D660 (2018).
122. M. L. Waskom, seaborn: Statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
123. J. J. A. Armenteros *et al.*, Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, e201900429 (2019).
124. S. K. Burley *et al.*, RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 (2021).
125. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
126. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. bioXriv [Preprint] (2022). https://doi.org/10.1101/2021.10.04.463034 (Accessed 4 October 2021).
127. A. Waterhouse *et al.*, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
128. B. Webb, A. Sali, Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinf.* **54**, 5.6.1–5.6.37 (2016).
129. G. M. Morris *et al.*, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).