nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-019-0355-x

A family of radical halogenases for the engineering of amino-acid-based products

Monica E. Neugebauer[1], Kiera H. Sumida[2], Jeffrey G. Pelton[3], Jonathan L. McMurry[2], Jorge A. Marchand[1] and Michelle C. Y. Chang[1,2,4,5]*

The integration of synthetic and biological catalysis enables new approaches to the synthesis of small molecules by combining the high selectivity of enzymes with the reaction diversity offered by synthetic chemistry. While organohalogens are valued for their bioactivity and utility as synthetic building blocks, only a handful of enzymes that carry out the regioselective halogenation of unactivated C_{sp^3}–H bonds have previously been identified. In this context, we report the structural characterization of BesD, a recently discovered radical halogenase from the Fe^{II}/α-ketogluturate-dependent family that chlorinates the free amino acid lysine. We also identify and characterize additional halogenases that produce mono- and dichlorinated, as well as brominated and azidated, amino acids. The substrate selectivity of this new family of radical halogenases takes advantage of the central role of amino acids in metabolism and enables engineering of biosynthetic pathways to afford a wide variety of compound classes, including heterocycles, diamines, α-keto acids and peptides.

The expansion of chemical diversity drives the discovery of small molecules and macromolecules with new function. In this regard, both synthetic and cellular chemistry provide access to an extremely broad range of compounds, but the structural space occupied by molecules made by humans as compared to those made by nature are often orthogonal. While living systems use the unparalleled selectivity of enzymes to construct molecules with a limited set of functional groups, synthetic methods instead utilize an extensive range of strategies for bond formation. As such, the development of approaches to bridge synthetic and cellular chemistry can help us to gain access to new structures and classes of compounds by combining the exquisite selectivity of enzymes with the breadth of functional groups used for synthetic transformations[1–3]. The introduction of halogens (X = F, Cl, Br, I) into a functional group-dense scaffold is especially useful, serving both to tune bioactivity[4,5] and to act as a reactive handle for the formation of new chemical bonds for diversification and modification of late-stage synthetic intermediates[6,7]. While the role of halogens in biosynthetic transformations is still being elucidated[8–11], their value in a variety of synthetic transformations and the design of synthetic routes is well accepted[12]. As a result, both synthetic and enzymatic approaches to the formation of carbon–halogen bonds are currently areas of major interest[6–8,13–15].

The selective modification of unactivated sp^3 C–H bonds is particularly difficult, as it is challenging to identify catalysts that are sufficiently powerful to activate these bonds while maintaining regio- and stereoselectivity[16]. Remarkably, nature has evolved a set of non-heme Fe^{II}/α-ketoglutarate (Fe^{II}/αKG)-dependent halogenase enzymes that can achieve this task, using a high-valent metal–oxo intermediate to generate a substrate radical that can rebound with the bound halide ligand[17–19]. However, only a handful of these radical halogenases have been characterized and are notable for their complex substrates[13]. The SyrB2 family has been found to maintain a strict requirement for carrier-protein-tethered substrates[20], whereas the WelO5 family halogenates late-stage indole alkaloid natural products[21]. As these types of intermediates are not readily modified using downstream enzymatic pathways, the discovery of new enzymes that act on simple and modular building blocks would greatly expand the biosynthetic potential of radical halogenation. We recently discovered a radical halogenase, BesD, that chlorinates the free amino acid lysine, making it the first of the αKG-dependent radical halogenases reported to chlorinate an amino acid without the requirement for a carrier protein[9].

In this work, we solved the crystal structure of BesD in complex with its substrate lysine to better understand the basis for direct halogenation of amino acids. We have also identified additional members of this new family of radical halogenases that chlorinate different polar and non-polar amino acids with varied regioselectivity by investigating the unexplored sequence space around the Fe^{II}/αKG-dependent halogenase, BesD. Additional active site mutagenesis and bioinformatic analysis provide further insight into the reaction mechanism of these enzymes and how they catalyze selective chlorination. Taking advantage of the central role of amino acid building blocks in the cellular production of metabolites, natural products and macromolecules[22–24], we demonstrate that the halogenated amino acid products can be further converted enzymatically to a range of compound classes such as nitrogen heterocycles, amines, keto acids and peptides. Taken together, this work greatly expands the utility of radical halogenases and highlights the promise of the BesD halogenase family in the biological production of diverse halogenated compounds.

Results

Structural characterization of BesD. The Fe^{II}/αKG-dependent halogenases are part of the large and highly diverse cupin superfamily, which contains members that catalyze a host of different reactions, including hydroxylation, halogenation, olefin epoxidation and stereoinversion on a broad range of substrates[25]. As such, the prediction and discovery of new activities using bioinformatic approaches has been limited by the sequence variability of family

[1]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. [3]QB3 Institute, University of California, Berkeley, Berkeley, CA, USA. [4]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [5]Lawrence Berkeley National Laboratory, Berkeley, CA, USA. *e-mail: mcchang@berkeley.edu

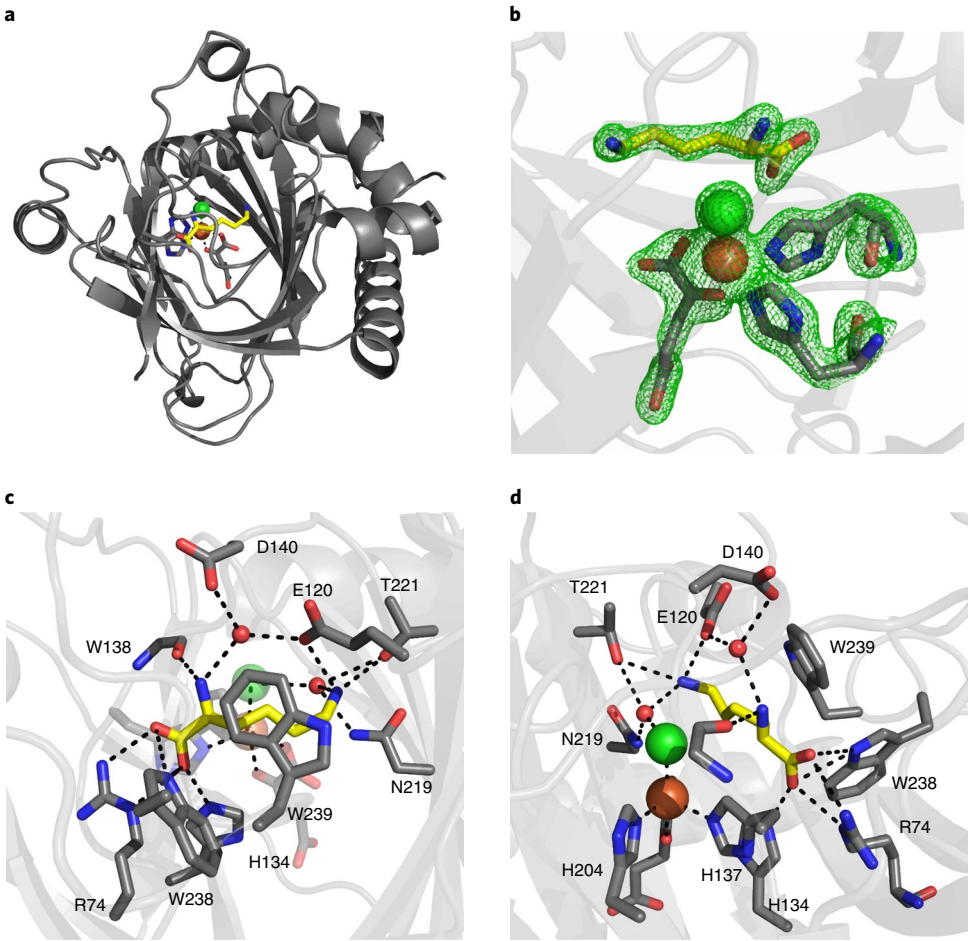

Fig. 1 | Crystal structure of lysine halogenase BesD. a, Overall fold of substrate-bound BesD. **b**, $F_o - F_c$ omit map for the lysine- (yellow) and Fe-bound complex with two histidines, Cl (green), Fe (brown) and αKG. The omit map was calculated by removing ligands before refinement and is contoured at ±3σ. **c**,**d**, Binding site of the substrate lysine from the front (**c**) and side (**d**). Hydrogen-bonding interactions coordinate the carboxylate and two amines of lysine. Trp239 stacks in front of the aliphatic side chain of lysine to cover the substrate in the active site.

members. In this case, BesD has very low identity to the carrier-protein-dependent halogenase SyrB2 (7% identity) and the indole alkaloid halogenase WelO5 (11% identity), and thus displays higher sequence similarity to predicted but uncharacterized Fe^{II}/αKG-dependent hydroxylases (≤46% identity) than to other known halogenases (Supplementary Fig. 1). A query of the superfamily database revealed that BesD groups in a different protein family within the cupin superfamily as compared to previously reported Fe^{II}/αKG-dependent halogenases[26].

To gain a better mechanistic understanding of BesD, we solved the X-ray crystal structure of BesD with lysine, Fe, chlorine and αKG bound in the active site (Fig. 1a,b and Supplementary Table 1). BesD was crystallized anaerobically with lysine and αKG, and soaked with Fe^{II}. The X-ray structure was solved at a resolution of 1.95 Å by Fe-SAD phasing. The structure contains four BesD monomers in the asymmetric unit, all of which possess the β-sandwich topology that is characteristic of Fe/αKG enzymes[18]. Each monomer contains a single Fe coordinated by His137, His204, chloride and αKG in a distorted square pyramidal geometry in the active site, which contains the HXG/A motif that is characteristic of halogenases. The binding site for lysine in BesD is largely polar, with a network of hydrogen bonds for binding to the carboxylate (Arg74, His134 and Trp238), α-amine (Asp140) and ε-amine (Glu120, Asn219 and Thr221) of the substrate (Fig. 1c,d). In addition, Trp239 on the C terminus of the protein stacks over the aliphatic side chain of lysine, closing over the substrate at the entrance of the active site. The structure also revealed a water network between Thr221 and the chloride ligand.

The crystal structure highlighted differences in substrate binding between BesD and the carrier-protein-dependent amino acid halogenases. In BesD, the positioning of the lysine substrate appears to be strongly determined by a two-point interaction between the free carboxylate of lysine and the guanidinium group of an active site arginine (Arg74). By contrast, in SyrB2, the carboxylate of the threonine substrate is ligated to a phosphopantetheine (Ppant) arm, which both tethers threonine to the carrier protein (SyrB1) through a thioester linkage and masks the carboxylate. It has been proposed that the Thr–Ppant arm enters the active site through a tunnel and that interactions with threonine and the Ppant arm, as well as SyrB1, are likely important for positioning the substrate in the active site for catalysis[27,28]. Another notable difference is that BesD appears to have a covering lid that holds lysine in the active site for catalysis. This substrate-binding mode is more reminiscent of the indole alkaloid halogenase, WelO5, in which the substrate is held into place by a helix that closes over the active site[19].

Catalytic selectivity in BesD. Understanding how radical halogenases catalyze halogenation selectively over the closely related hydroxylation reaction remains an active area of investigation. Fe^{II}/αKG-dependent halogenases are closely related both evolutionarily and mechanistically to the larger and more well-studied class of Fe^{II}/αKG-dependent hydroxylases[29,30]. Upon activation with oxygen

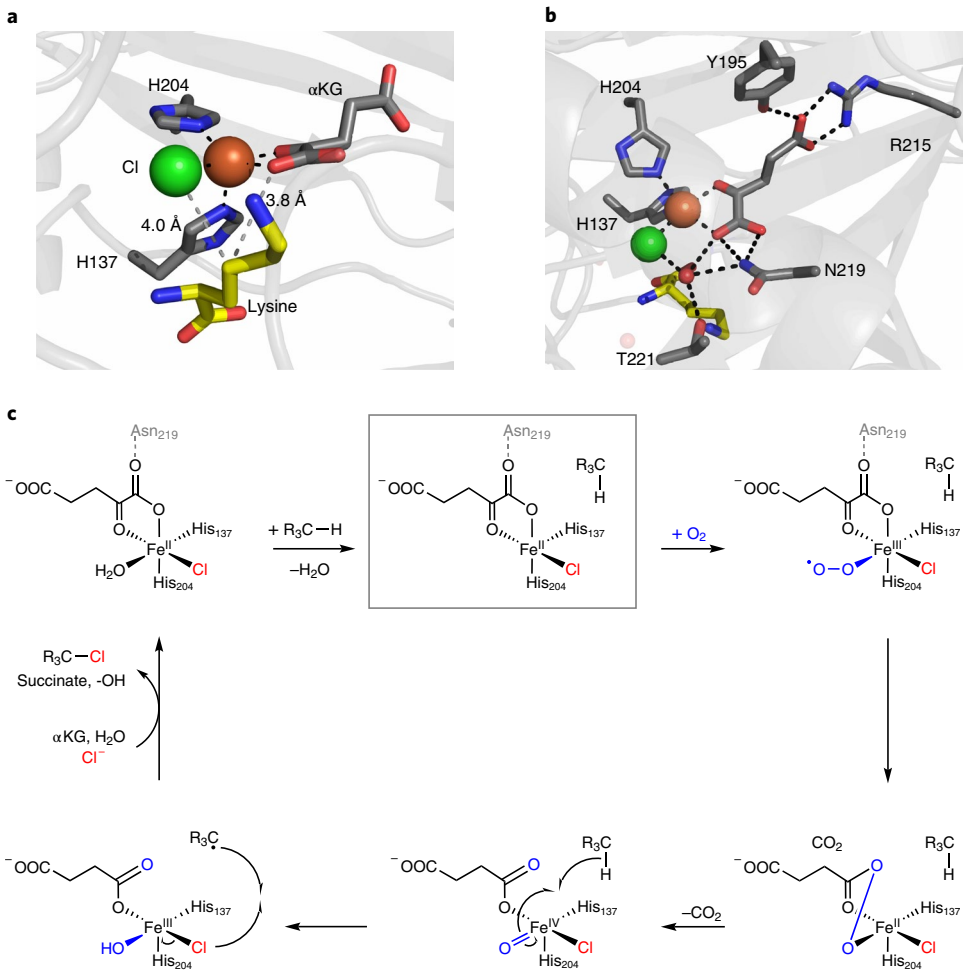

Fig. 2 | Proposed mechanism of halogenation by BesD. a, The ligand complex displays a vacant site for oxygen binding that is distal to the substrate. The target γ-carbon of lysine is 4.0 Å from Cl and 3.8 Å from αKG (indicated with gray dotted lines). **b**, Top view of the Fe site showing αKG contacts and hydrogen bonding between the C1-carboxylate of αKG and Asn219. **c**, Proposed mechanism of halogenation by BesD. The Fe is coordinated by His137, His204, chloride and αKG in a distorted square pyramidal geometry (gray box). On the basis of the crystal structure in this study, the putative vacant site for oxygen binding appears to be *trans* to His137, on the opposite side of the His204–Cl–αKG plane from the substrate carbon that is targeted for hydrogen atom abstraction. In this class of enzymes, binding of the substrate followed by oxygen leads to decarboxylation of αKG via attack of the distal oxygen on C2 of αKG. It is possible that conformational change of the substrate or a shift of the oxo ligand toward the position *trans* to His204 before, during, or after decarboxylation of aKG may be needed to facilitate hydrogen atom abstraction. Finally, rebound of the suitably positioned halogen would yield the chlorinated product. Note that the precise positions of the oxo species in the proposed intermediates are unknown.

and αKG, both hydroxylases and halogenases generate a high-valent Fe^{IV}–oxo species, which can abstract a hydrogen atom from the substrate[31]. For hydroxylases, rebound of the substrate radical with the resulting hydroxo group is the only possible pathway, as the coordination site occupied by a halide in halogenases is filled by the Asp or Glu ligand from the HXD/E motif. In halogenases, however, rebound with either the halogen or the hydroxyl radical is possible, leading to potential reaction partitioning between halogenation and hydroxylation, respectively[32] (Supplementary Fig. 2). In the case of BesD, substitution of the active site glycine of the HXG motif to an aspartate eliminates halogenation activity and results in only hydroxylation[9]. Indeed, it has been shown that halogenases can carry out a low level of off-pathway hydroxylation, especially when challenged with non-native substrates[33]. Efforts to engineer halogenases from hydroxylases have also been difficult, suggesting that second-sphere interactions are important in controlling reaction partitioning[34].

Combining insights from our structure with previous studies of Fe^{II}/αKG-dependent enzymes[17], we can propose a mechanism for halogenation by BesD. Halogenases must orient the Fe^{IV}–oxo intermediate such that it is suitably positioned to perform hydrogen atom abstraction while disfavoring rebound of the resulting Fe^{III}–OH intermediate in favor of Fe^{III}–Cl rebound. Our structure captured under anaerobic conditions most likely represents a snapshot of the active site before oxygen binding. By analogy to other Fe^{II}/αKG-dependent enzymes in the literature[18,19], the putative vacant site for oxygen binding appears to be *trans* to His137, on the opposite side of the His204–Cl–αKG plane from the substrate carbon that is targeted for hydrogen atom abstraction (Fig. 2a,b and Supplementary Fig. 3). If the oxo moiety remains *trans* to His137 upon formation of the Fe^{IV}–oxo intermediate, and if the substrate position does not change, the distance between the oxo moiety and site of hydrogen atom abstraction (C4 of lysine) would be ~5.0 Å on the basis of the Fe^{II}–C4(lysine) distance (4.8 Å) and the estimated Fe^{IV}–oxo bond length (1.6 Å)[31]. While the precise structures of the reaction intermediates are unknown, this distance is longer than that calculated from density-functional theory studies on SyrB2 (ref. [31]). It is also possible that conformational change of the substrate or a shift[19,35] of the oxo ligand toward the position *trans* to His204 before, during or after decarboxylation of αKG may be needed to facilitate hydrogen

atom abstraction to bring the substrate and Fe-center geometries within the required range[31]. Finally, rebound of a suitably positioned halogen would yield the chlorinated product (Fig. 2c).

To further investigate catalytic selectivity in BesD, we performed alanine scan mutagenesis of active site residues in BesD. Protein variants were expressed, purified and incubated with lysine before product analysis by LC–MS (Supplementary Fig. 4). Most alanine variants abolished all activity (both halogenation and hydroxylation). However, three variants yielded interesting profiles (Fig. 3). Strikingly, H134A increased overall enzyme activity and abolished enzyme selectivity, resulting in decreased halogenation to hydroxylation ratios. The N219A variant also abolished selectivity for halogenation, albeit with reduced activity as compared to wild-type BesD. Interestingly, the T221A substitution had no effect on activity or halogenation selectivity despite the conservation of a polar amino acid at position 221 in halogenases (Fig. 3c).

Analysis of the BesD crystal structure in conjunction with sequence conservation of these residues suggested that His134 and Asn219 play important roles in the second-sphere interactions that control reaction partitioning. His134, which is highly conserved, is hydrogen-bonded to the carboxylate of lysine in the BesD structure and thus presumably plays an important role in orienting lysine relative to the iron complex. The loss in halogenation selectivity upon mutation of His134 is consistent with studies that implicate the precise positioning of the substrate within the active site to favor halogen rebound over hydroxyl rebound[33,36]. At 2.9 Å away from αKG, Asn219 may serve to orient αKG through hydrogen bonding such that the vacant putative oxygen-binding site is located far from the substrate. Additionally, Asn219 could be involved in orienting the Fe^{III}–OH intermediate to differentially favor halogen rebound, analogous to the proposed role of Ser189 in WelO5[19], if the oxo ligand does shift toward the substrate to facilitate hydrogen atom abstraction (Supplementary Fig. 3). Upon examining the sequence conservation of key active site residues in predicted halogenase and hydroxylase sequences with homology to BesD, we found that Asn219 was highly conserved in the halogenase homologs while greater variability was displayed by hydroxylase homologs (Supplementary Fig. 5). This conservation of Asn219 further supports the importance of second-sphere interactions in the BesD amino acid halogenase family.

Discovery of new amino acid halogenases. To explore the sequence space around BesD, we performed a sequence-based homology search (BLAST E-value of e^{-5}) to identify other potential halogenase candidates. The related hydroxylases were filtered from the dataset on the basis of the presence of the characteristic HXD/E motif, as halogenases are known to possess an HXG/A motif instead to allow for halide coordination to Fe during catalysis[18]. To maximize sampling of functional diversity, we generated a sequence similarity network (SSN)[37] to group the homologous enzymes into eight new putative isofunctional groups, which we termed clusters A–H (Supplementary Figs. 6 and 7, and Supplementary Data). Genes adjacent to putative halogenase coding sequences were analyzed with the Enzyme Function Initiative's Genome Neighborhood Tools to identify conserved genomic contexts associated with these enzymes[37]. This analysis revealed numerous genes encoding proteins predicted to be involved in amino acid metabolism, such as amino acid transporters, LysR family transcriptional regulators, serine hydroxymethyl transferases, and ATP-grasp-dependent carboxy-amine ligases (Supplementary Fig. 8). The conservation

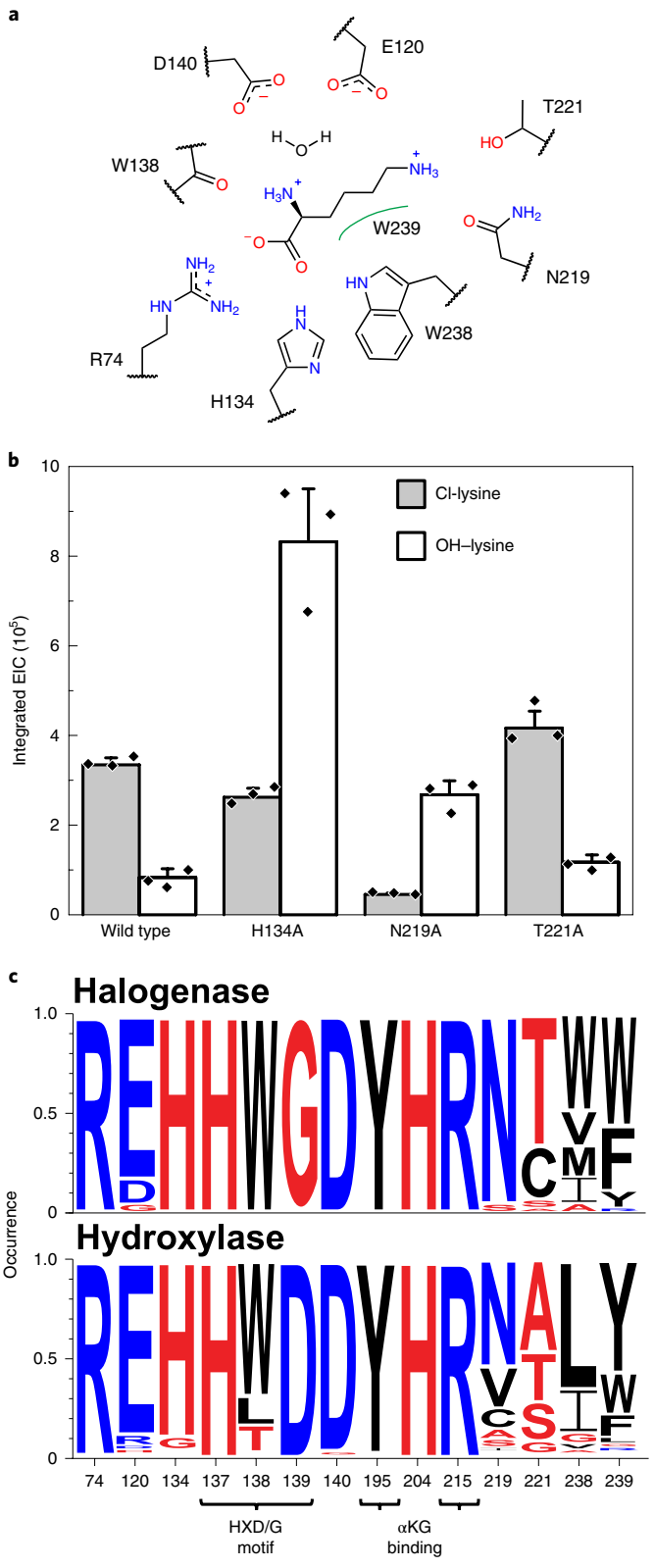

Fig. 3 | Alanine scan of active site residues. a, Lysine-binding residues of BesD. **b**, LC–MS analysis of reaction products following incubation of wild-type or mutant SIBesD enzymes with Fe^{II}, αKG, NaCl and lysine. Integrated extracted ion chromatograms (EICs) for the formation of Cl-lysine (gray) and hydroxylysine (white) by wild-type, H134A, N219A and T221A SIBesD enzymes. R74A, D140A, E120A, W238A and W239A variants yielded neither product (Supplementary Fig. 4). Data are mean ± s.d. from three technical replicates. **c**, LOGOS plot of sequence conservation of active site residues within halogenase (top) and hydroxylase (bottom) homologs of BesD. Homologs of BesD (BLAST E-value of e^{-5}) were sorted on the basis of the presence of the HXG (halogenase) or HXD (hydroxylase) motifs.

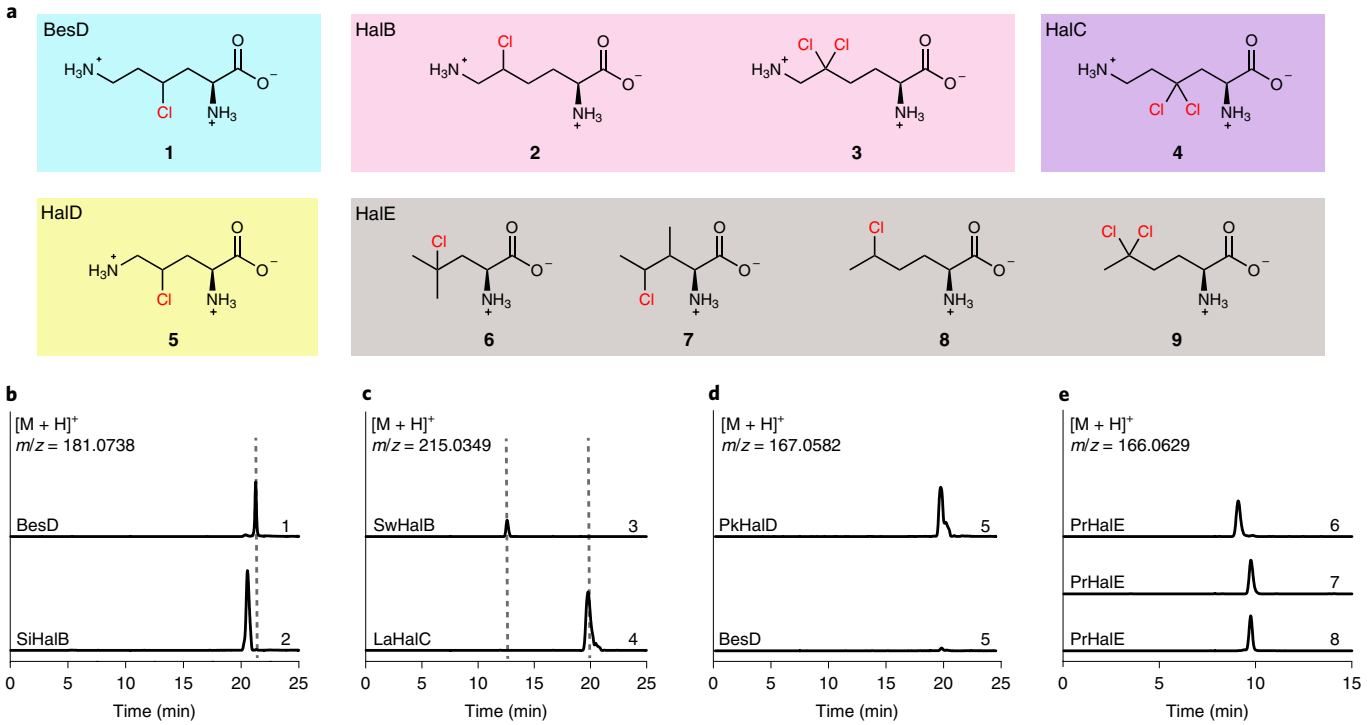

Fig. 4 | Amino acid halogenase diversity. a, BesD homologs and products of enzymatic amino acid halogenation. **b**, Regioselective monohalogenation is observed with 4-Cl-lysine (**1**, $m/z = 181.0738$) or 5-Cl-lysine (**2**, $m/z = 181.0738$) produced by BesD or SiHalB, respectively. **c**, Regioselective dihalogenation is also observed with 5,5-dichlorolysine (**3**, $m/z = 215.0349$) or 4,4-dichlorolysine (**4**, $m/z = 215.0349$) produced by SwHalB or LaHalC, respectively. **d**, Ornithine is a substrate for both PkHalD and BesD (4-Cl-ornithine, **5**, $m/z = 167.0582$). **e**, The non-polar amino acids leucine, isoleucine and norleucine are substrates for PrHalE (4-Cl-leucine, **6**, $m/z = 166.0629$; 4-Cl-isoleucine, **7**, $m/z = 166.0629$; 5-Cl-norleucine, **8**, $m/z = 166.0629$). Peak intensities are normalized in **e**. Assays include Fe^{II}, ascorbate, αKG and NaCl, in addition to the amino acid substrate. All EICs are representative of at least three experimental replicates.

of genes encoding proteins associated with amino acid metabolism suggested that the BesD homologs might also be amino acid halogenases. We cloned, heterologously expressed and purified representative halogenase homologs from clusters A–H and tested the proteins for activity against a panel of amino acids (Supplementary Fig. 9). The products were analyzed by LC–MS, revealing newly accessible chlorinated amino acid products and identifying substrates for enzymes within the clusters (Fig. 4a and Supplementary Tables 2–4).

Among the characterized halogenases, we discovered members with different regioselectivities. BesD and enzymes from cluster A halogenate lysine at the γ-carbon to produce 4-Cl-lysine (**1**). However, HalB from *Streptomyces iranensis* (SiHalB) yielded a product with the same exact mass and characteristic chlorinated isotope pattern as 4-Cl-lysine, but with a different retention time (Fig. 4b and Supplementary Fig. 10). Through NMR analysis of the methyl esterified product, we identified the product of SiHalB as 5-Cl-lysine (**2**; Supplementary Note 1). Although they are 43% identical, SiHalB and BesD have distinct regioselectivities, highlighting the capability of Fe^{II}/αKG-dependent enzymes to perform regioselective halogenation under mild conditions. We also found halogenases that perform both mono- and dichlorination of lysine (Fig. 4c). When lysine was incubated with BesD, only monochlorinated 4-Cl-lysine was observed. However, in addition to monochlorination of lysine, *Legionella anisa* HalC (LaHalC) could also carry out dichlorination to produce 4,4-dichlorolysine (**4**; Supplementary Fig. 11 and Supplementary Note 1) while HalB from *Streptomyces wuyuanensis* (SwHalB) yielded 5,5-dichlorolysine (**3**; Supplementary Fig. 12 and Supplementary Note 1). Dichlorination has been observed in radical carrier-protein-dependent halogenases as well[13].

In addition to differing regioselectivities, other BesD family members were identified that could act on other amino acid substrates. Enzymes from cluster D exhibited chain-length preference and halogenated the five-carbon substrate ornithine preferentially over the six-carbon substrate lysine. In particular, the *Pseudomonas kilonensis* HalD (PkHalD) was found to prefer ornithine ($k_{cat}/K_M = 330 \pm 70\ mM^{-1}min^{-1}$) as a substrate by 25-fold over lysine ($k_{cat}/K_M = 13.2 \pm 3.6\ mM^{-1}min^{-1}$) (Fig. 4d, Supplementary Figs. 13 and 14, and Supplementary Note 1). Finally, cluster E was found to contain aliphatic amino acid halogenases. When we assayed *P.* sp. *Root562* HalE (PrHalE), we observed no activity on the amino acids lysine and ornithine. However, when we incubated PrHalE with the aliphatic amino acids leucine, isoleucine and norleucine, we observed the expected mass and isotopic patterns for chlorinated products of those amino acids (**6**–**8**; Fig. 4e, Supplementary Figs. 15 and 16, and Supplementary Note 1). We additionally observed a dichlorinated species (**9**) when norleucine was used as a substrate. Further characterization by NMR indicated that PrHalE chlorinates leucine and isoleucine (CH_2) at the γ-position and norleucine at the δ-position (Supplementary Note 1).

Remarkably, halogenases from clusters A–H have evolved to accommodate different amino acids, with regioselectivity for specific sites on those substrates, while maintaining fidelity for halogenation over the competing side reaction of hydroxylation. To further explore the structural basis for these differences, the sequences of halogenases A–E were aligned (Supplementary Fig. 17). As expected, the carboxylate- and amine-binding residues Arg74, His134 and Asp140 are highly conserved, regardless of substrate preference. However, it was surprising to note that the ε-amine-binding residues of BesD (Glu120, Asn219 and Thr221) were

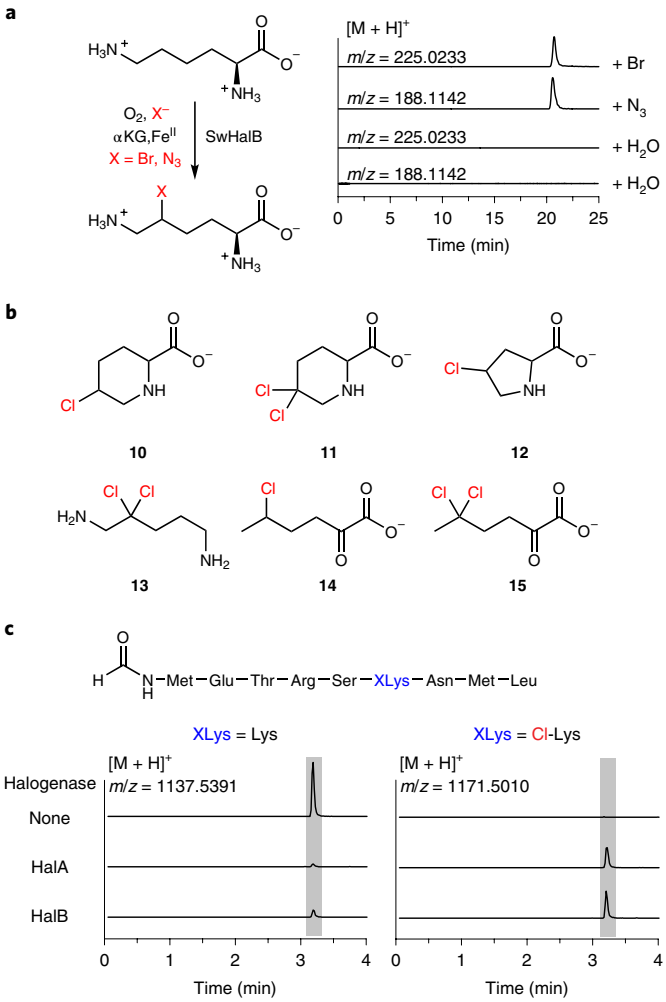

Fig. 5 | Engineering downstream pathways with amino acid halogenases. **a**, Bromination and azidation can be carried out with BesD halogenases. Bromolysine ($m/z = 225.0233$) and azidolysine ($m/z = 188.1142$) are produced by SwHalB using NaBr or NaN_3, respectively. EICs are representative of at least three experimental replicates. **b**, The incubation of substrates (lysine, ornithine or norleucine) and cofactors (Fe^{II}, αKG and NaCl) with various halogenases and amino-acid-metabolizing enzymes yields chlorinated heterocycles (**10**–**12**), chlorinated diamines (**13**) and chlorinated α-keto acids (**14** and **15**). Compounds **10** and **11** are the products of incubating lysine with SwHalB and lysine cyclodeaminase; compound **12** is the product of incubating ornithine with PkHalD and ornithine cyclodeaminase; compound **13** is the product of incubating lysine with SwHalB and lysine decarboxylase; and compounds **14** and **15** are the products of incubating norleucine with PrHalE and the aliphatic amino acid transaminase, IlvE. At least three experimental replicates were carried out. **c**, PfHalA or SwHalB could be used to generate Cl-lysine in situ followed by IVTT initiated by addition of PURExpress kit components along with a plasmid encoding the METRSKNML peptide. The EIC shows the production of peptides containing lysine ($m/z = 1137.5391$) or Cl-lysine ($m/z = 1171.5010$). Data are representative of three experimental replicates.

also highly conserved, even for halogenases that act on non-polar amino acids. The greatest disparity in sequence occurs at Trp238 and Trp239, which are located at the C-terminal end of the protein. Sequence alignment revealed that the C-terminal regions of halogenases A–E are highly variable, differing by as many as ten amino acids in length and resulting in unreliable alignments of that region. Structural analysis of BesD showed that the C terminus closes over the substrate in the active site of the protein, occluding it from solvent (Supplementary Fig. 18). The C terminus of BesD additionally interacts with a corresponding internal loop near the active site (residues 117–119), which also varies among the halogenase clusters. Combined with the structure, the sequence analysis suggested that the C terminus and the corresponding loop could be important determinants of substrate selectivity.

Biosynthetic applications of amino acid halogenases. Amino acids are key building blocks for metabolism, serving as the source for many metabolites and natural products, as well as ribosomally and non-ribosomally synthesized peptides and proteins. Therefore, the discovery of the BesD family of halogenases raises the possibility that halogens and other functional groups could be incorporated into these classes of compounds through engineering of downstream biosynthetic transformations. We first examined the ability to expand the repertoire of the amino acid halogenases to carry out bromination and azidation by replacement of the Cl^- ligand with Br^- and N_3^-, respectively[38]. When NaCl was replaced with NaBr and NaN_3, SwHalB catalyzed the corresponding bromination and azidation reactions as observed by LC–MS (Fig. 5a). Thus, a variety of functional handles can be installed onto amino acid substrates to diversify amino acid building blocks using SwHalB simply by providing an alternate ligand for the Fe complex. In particular, bromine and azide functional groups serve as handles for substitution reactions[39] and as precursors to amines[40] and nitrogen-containing heterocycles[41], respectively.

Amino acids also participate in a broad range of metabolic pathways and can be directly transformed into a variety of new compound classes (Fig. 5b). For example, heterocycles are key structural components in a number of bioactive molecules and drugs[42] and can be made by oxidative cyclization by lysine cyclodeaminase (RapL)[43] and ornithine cyclodeaminase (OCD)[44] to form pipecolate and proline, respectively. To test whether the cyclodeaminase enzymes were sufficiently promiscuous to generate chlorinated heterocycles, we cloned, expressed and purified RapL and OCD (Supplementary Fig. 19). Coupled reactions with lysine or ornithine halogenases and RapL or OCD, respectively, yielded chlorinated compounds with the predicted masses and isotope patterns of chloropipecolate (**10**), dichloropipecolate (**11**) and chloroproline (**12**) (Supplementary Fig. 20 and 21). Thus, the newly discovered amino acid halogenases can be further enzymatically elaborated to form five- and six-membered chlorinated heterocycles.

Owing to the utility of diamines as precursors for polymers such as nylon[45], we next sought to access chlorinated diamines through coupled enzymatic reactions. Lysine decarboxylase (LDC) catalyzes the pyrodoxal-5′-phosphate (PLP)-dependent decarboxylation of lysine to 1,5-pentanediamine[46]. LDC from *Escherichia coli* was heterologously expressed and purified to evaluate its tolerance of a chlorinated substrate. To generate a chlorinated diamine, LDC was added to a reaction following the enzymatic synthesis of 5,5-dichlorolysine by SwHalB. A mass consistent with a dichlorinated diamine (**13**) was observed by LC–MS (Supplementary Fig. 22). We additionally accessed chlorinated α-keto acids through IlvE, an aliphatic amino acid aminotransferase. IlvE from *E. coli* has been shown to catalyze transamination of leucine, isoleucine and norleucine into their corresponding α-keto acids[47]. Incubation of enzymatically synthesized chlorononorleucine with IlvE yielded products with the predicted mass and isotope pattern of chlorinated α-keto acids (**14** and **15**) as observed by LC–MS (Supplementary Fig. 23).

Finally, we investigated whether chlorolysine synthesized by BesD homologs could be incorporated into peptides with an in vitro transcription and translation (IVTT) system[48]. Following incubation of amino acids with PfHalA or SwHalB, transcription and translational machinery from the PURExpress IVTT kit, along with a plasmid encoding a nine-amino-acid peptide, were added to the

reaction mixture. When the amino acids were preincubated with either halogenase, a product corresponding to the exact mass of the chlorinated peptide was detected by LC–MS (Fig. 5c). Therefore, the halogenases from the BesD family can be coupled with IVTT systems to generate chlorinated peptides from 4-Cl-lysine and 5-Cl-lysine.

Discussion

Enzymes serve as highly useful catalysts, whose ability to achieve high selectivity and turnover numbers under mild conditions have advanced the development of streamlined syntheses of a broad range of compounds[3]. One particularly interesting reaction carried out by enzymes is the selective activation of C–H bonds to replace them with halogen substituents, which enables the functionalization of small molecules as semisynthetic intermediates for a broad range of downstream reaction pathways or to tune the bioactivity of lead compounds. The discovery of Fe^{II}/αKG-dependent halogenases that can operate directly on simple amino acids expands the scope of known radical halogenation chemistry. Furthermore, this halogenase family opens the door to the biosynthetic engineering of a large number of small molecule and macromolecular targets derived from amino acid building blocks.

Characterization of the radical halogenases in the BesD family has shown that these enzymes accept both polar and non-polar amino acids, demonstrate regioselective halogenation of C_{sp^3}–H bonds and carry out both mono- and dichlorination. Structural studies of BesD further provide valuable insight into its substrate selectivity and the potential mechanisms by which reaction partitioning between halogenation and hydroxylation is achieved. In addition to showing that bromination and azidation are catalyzed by the BesD family, we have demonstrated that downstream enzymatic pathways can be constructed to utilize halogenase-modified intermediates to produce different classes of compounds, including nitrogen heterocycles, diamines, keto acids and ribosomally synthesized peptides. Altogether, these results highlight the potential to integrate this family of halogenases directly into engineered biological reaction pathways and expand our ability to access a wide variety of new compounds.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41589-019-0355-x.

Received: 16 January 2019; Accepted: 31 July 2019;
Published online: 23 September 2019

References

1. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409**, 232–240 (2001).
2. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132 (2017).
3. Savile, C. K. et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
4. Harris, C., Kannan, R., Kopecka, H. & Harris, T. The role of the chlorine substituent in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *J. Am. Chem. Soc.* **107**, 6652–6658 (1985).
5. Groll, M., Huber, R. & Potts, B. C. M. Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of β-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **128**, 5136–5141 (2006).
6. Latham, J., Brandenburger, E., Shepherd, S. A., Menon, B. R. K. & Micklefield, J. Development of halogenase enzymes for use in synthesis. *Chem. Rev.* **118**, 232–269 (2018).
7. Gkotsi, D. S., Dhaliwal, J., McLachlan, M. M., Mulholand, K. R. & Goss, R. J. Halogenases: powerful tools for biocatalysis (mechanisms applications and scope). *Curr. Opin. Chem. Biol.* **43**, 119–126 (2018).
8. Neumann, C. S., Fujimori, D. G. & Walsh, C. T. Halogenation strategies in natural product biosynthesis. *Chem. Biol.* **15**, 99–109 (2008).
9. Marchand, J. A. et al. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019).
10. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005).
11. Nakamura, H., Schultz, E. E. & Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **13**, 916–921 (2017).
12. Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry* (University Science, 2006).
13. Agarwal, V. et al. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017).
14. Liang, T., Neumann, C. N. & Ritter, T. Introduction of fluorine and fluorine-containing functional groups. *Ange. Chem. Int. Ed. Engl.* **52**, 8214–8264 (2013).
15. Petrone, D. A., Ye, J. & Lautens, M. Modern transition-metal-catalyzed carbon–halogen bond formation. *Chem. Rev.* **116**, 8003–8104 (2016).
16. Shilov, A. E. & Shul'pin, G. B. Activation of C–H bonds by metal complexes. *Chem. Rev.* **97**, 2879–2932 (1997).
17. Bollinger, J. M. et al. in *2-Oxoglutarate-Dependent Oxygenases* (eds Hausinger R.P. & Schofield, C.J.) 95–122 (Royal Society of Chemistry, London, 2015).
18. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006).
19. Mitchell, A. J. et al. Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016).
20. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe^{II} α-ketoglutarate- and O_2-dependent halogenase. *Proc. Natl Acad. Sci. USA* **102**, 10111–10116 (2005).
21. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 6–10 (2014).
22. Ortega, M. A. & van der Donk, W. A. New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol.* **23**, 31–44 (2016).
23. Runguphan, W., Qu, X. & O'Connor, S. E. Integrating carbon–halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461–464 (2010).
24. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
25. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).
26. Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B. & Gough, J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* **47**, D490–D494 (2019).
27. Kulik, H. J. & Drennan, C. L. Substrate placement influences reactivity in non-heme Fe(II) halogenases and hydroxylases. *J. Biol. Chem.* **288**, 11233–11241 (2013).
28. Matthews, M. L. et al. Substrate-triggered formation and remarkable stability of the C–H bond-cleaving chloroferryl intermediate in the aliphatic halogenase, SyrB2. *Biochemistry* **48**, 4331–4343 (2009).
29. Puri, M., Biswas, A. N., Fan, R., Guo, Y. & Que, L. Modeling non-heme iron halogenases: high-spin oxoiron(IV)–halide complexes that halogenate C–H bonds. *J. Am. Chem. Soc.* **138**, 2484–2487 (2016).
30. Galonić, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* **3**, 113–116 (2007).
31. Wong, S. D. et al. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013).
32. Srnec, M. & Solomon, E. I. Frontier molecular orbital contributions to chlorination versus hydroxylation selectivity in the non-heme iron halogenase SyrB2. *J. Am. Chem. Soc.* **139**, 2396–2407 (2017).
33. Matthews, M. L. et al. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl Acad. Sci. USA* **106**, 17723–17728 (2009).
34. Mitchell, A. J. et al. Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate. *Biochemistry* **56**, 441–444 (2017).
35. Zhang, Z. et al. Crystal structure of a clavaminate synthase–Fe(II)–2-oxoglutarate–substrate–NO complex: evidence for metal centered rearrangements. *FEBS Lett.* **517**, 7–12 (2002).
36. Martinie, R. J. et al. Experimental correlation of substrate position with reaction outcome in the aliphatic halogenase, SyrB2. *J. Am. Chem. Soc.* **137**, 6912–6919 (2015).
37. Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* **56**, 4293–4308 (2017).
38. Matthews, M. L. et al. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014).
39. Fu, G. C. Transition-metal catalysis of nucleophilic substitution reactions: a radical alternative to S_N1 and S_N2 processes. *ACS Cent. Sci.* **3**, 692–700 (2017).

40. Nyffeler, P. T., Liang, C.-H., Koeller, K. M. & Wong, C.-H. The chemistry of amine–azide interconversion: catalytic diazotransfer and regioselective azide reduction. *J. Am. Chem. Soc.* **124**, 10773–10778 (2002).
41. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48**, 6974–6998 (2009).
42. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011).
43. Gatto, G. J., Boyne, M. T., Kelleher, N. L. & Walsh, C. T. Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster. *J. Am. Chem. Soc.* **128**, 3838–3847 (2006).
44. Goodman, J. L. et al. Ornithine cyclodeaminase: structure, mechanism of action, and implications for the μ-crystallin family. *Biochemistry* **43**, 13883–13891 (2004).
45. Wendisch, V. F., Mindt, M. & Pérez-García, F. Biotechnological production of mono- and diamines using bacteria: recent progress, applications, and perspectives. *Appl. Microbiol. Biotechnol.* **102**, 3583–3594 (2018).
46. Takatsuka, Y., Yamaguchi, Y., Ono, M. & Kamio, Y. Gene cloning and molecular characterization of lysine decarboxylase from *Selenomonas ruminantium* delineate its evolutionary relationship to ornithine decarboxylases from eukaryotes. *J. Bacteriol.* **182**, 6732–6741 (2000).
47. Rudman, D. & Meister, A. Transamination in *Escherichia coli*. *J. Biol. Chem.* **200**, 591–604 (1953).
48. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).

Acknowledgements

This work received support from the National Science Foundation (CHE-1710588) and the Department of Energy (DOE/LBNL DEAC02-05CH11231, FWP CH030201). M.E.N. acknowledges the support of a National Science Foundation graduate research fellowship. J.L.M. acknowledges the support of a National Institutes of Health NRSA training grant (1 T32 GMO66698). J.A.M. acknowledges the support of a University of California, Berkeley Chancellor's fellowship, Howard Hughes Medical Institute Gilliam fellowship and National Institutes of Health NRSA training grant (1 T32 GMO66698). X-ray data were collected at the Advanced Light Source Beamline 8.3.1, which is operated by the University of California Office of the President, Multicampus Research Programs and Initiatives (MR-15-328599), the National Institutes of Health (R01 GM124149 and P30 GM124169), Plexxikon and the Integrated Diffraction Analysis Technologies program of the U.S. Department of Energy Office of Biological and Environmental Research. The Advanced Light Source is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the U.S. Department of Energy under contract number DEAC02-05CH11231, Office of Basic Energy Sciences. The funds for the 900-MHz NMR spectrometer housed in the QB3 Institute in Stanley Hall at University of California, Berkeley were provided by the National Institutes of Health (GM68933). We thank E. C. Wittenborn, J. Holton, C. Gee and G. Meigs for crystallography advice. We also thank J.M. Bollinger and A.K. Boal for helpful discussions.

Author contributions

M.E.N. carried out protein crystallography, bioinformatics and enzyme characterization experiments. K.H.S. carried out enzyme characterization experiments. J.G.P. performed NMR experiments. J.L.M. contributed to bioinformatics. J.A.M. contributed helpful discussions and contributed to bioinformatics. M.E.N., M.C.Y.C. and K.H.S. planned experiments. M.E.N. and M.C.Y.C. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-019-0355-x.

Correspondence and requests for materials should be addressed to M.C.Y.C.

Reprints and permissions information is available at www.nature.com/reprints.

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

Methods

Commercial materials. LB medium Miller, LB agar Miller, Terrific broth (TB) and glycerol were purchased from EMD Biosciences. Carbenicillin, kanamycin, chloramphenicol, IPTG, sodium chloride, DTT, HEPES, magnesium chloride hexahydrate, acetonitrile, EDTA, hydrochloric acid, 3-kDa molecular weight cut-off (MWCO) dialysis tubing, ammonium bicarbonate, sodium acetate, chloroform and sodium hydroxide were purchased from Thermo Fisher Scientific. Phosphoenolpyruvate (PEP), ATP sodium salt, NADH, pyruvate kinase, lactate dehydrogenase, lysozyme, poly(ethyleneimine) solution, ammonium iron(II) sulfate hexahydrate, α-ketoglutaric acid sodium salt, β-mercaptoethanol (βME), *N*,*N*,*N*′,*N*′-tetramethyl-ethane-1,2-diamine (TEMED), betaine, dimethylsulfoxide (DMSO), PLP, sodium L-ascorbate, acetonitrile (LC–MS-grade), ammonium formate (LC–MS-grade), deuterium oxide, L-lysine for crystallography, 20 canonical amino acids, seleno-L-methionine, L-leucine ($^{13}C_6$, 98%), L-isoleucine ($^{13}C_6$, 98%; ^{15}N, 98%), DL-norleucine, 2-(*N*-morpholino)ethanesulfonic acid hydrate, 4-morpholineethanesulfonic acid (MES), Pall Nanosep centrifugal Omega membrane 3K, sodium bromide and sodium azide were purchased from Sigma-Aldrich. PageRuler Plus prestained protein ladder was purchased from Thermo Fisher Scientific. Succinyl-CoA synthetase was purchased from Megazyme International. Methanol (D_4, 99.8%), chloroform (D, 99.8%), L-lysine·2HCl ($^{13}C_6$, 99%; $^{15}N_2$, 99%) and L-ornithine ($^{13}C_5$, 99%; $^{15}N_2$, 99%) were purchased from Cambridge Isotope Laboratories. Formic acid was purchased from Acros Organics. Restriction enzymes, T4 DNA ligase, Antarctic phosphatase, Phusion DNA polymerase, Q5 DNA polymerase, T5 exonuclease, Taq DNA ligase and the PURExpress IVTT kit Δ(aa, tRNA) were purchased from New England Biolabs. Deoxynucleotides (dNTPs), were purchased from Invitrogen. Oligonucleotides were purchased from Integrated DNA Technologies, resuspended at a stock concentration of 100 μM in water and stored at either 4 °C for immediate usage or −20 °C for long-term storage. DNA-purification kits and Ni-NTA agarose were purchased from Qiagen. Zirconia–silica beads were purchased from BioSpec Products. Complete EDTA-free protease inhibitor was purchased from Roche Applied Science. PD-10 desalting columns and Superdex 75 16/60-pg columns were purchased from GE Healthcare. Amicon Ultra 10,000 MWCO centrifugal concentrators and Milli-Q Gradient water purification system were purchased from Millipore. Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis-grade SDS and ammonium persulfate were purchased from Bio-Rad Laboratories. Ultrayield baffled flasks were purchased from Thompson Instrument Company.

Bioinformatic selection of BesD halogenase homologs for screening. To select halogenase homologs for functional screening, the sequence of BesD (WP_014151497) was used as a query for BLAST against the non-redundant protein database with a cutoff value of e^{-5}, yielding 261 hits. BLAST hits were aligned using MUSCLE[49] and visualized in AliView[50], and the sequence position of the HXD/E or HXG/A motif was identified manually. A Biopython script was used to parse the list of alignments to segregate the halogenases (HXG/A) for storage in a separate file, yielding 105 sequences, which were retrieved using BatchEntrez (NCBI). From the list of halogenases, redundant sequences of 98% identity were removed using CD-HIT[51]. The remaining halogenase sequences with homology to BesD were input into the EFI-EST web tool[37] (https://efi.igb.illinois.edu/efi-est), using option C to read FASTA files with headers. After initial processing, sequences below 150 amino acids in length were excluded from the sequence similarity network. Using Cytoscape v.3.6.1, the network was adjusted by deleting edges with low alignment scores until the halogenases identified for Bes biosynthesis[9] were separated into an isofunctional cluster, which occurred at an alignment score value of 88, which corresponds to a sequence identity of ~56%. The EFI-GNT web tool[37] (https://efi.igb.illinois.edu/efi-gnt) was used to identify adjacent genes that were shared among 50% of the halogenase sequences in each cluster, revealing genes associated with amino acid metabolism (Supplementary Fig. 8).

Phylogenetic analysis of BesD homologs. The list of halogenases for phylogenetic analysis was obtained as described above. Halogenase sequences were aligned using MUSCLE. The MEGAX interface was then used to analyze the alignment by maximum likelihood on the basis of the JTT matrix-based model[52], and uncertainty in the topology of the resulting tree was evaluated with 500 bootstrap replicates.

Bioinformatic analysis of active site conservation among halogenases versus hydroxylases. Separate lists of halogenases (HXG/A motif) and hydroxylases (HXD/E motif) were obtained as described above. Redundant sequences with at least 90% identity to other sequences under consideration were removed using CD-HIT, leaving 48 halogenases and 97 hydroxylases. Next, the halogenase and hydroxylase lists were separately analyzed with LOGOS[53] (http://weblogo.threeplusone.com/create.cgi) to generate a graphical representation of sequence conservation within each dataset.

Bioinformatic analysis of the variability of the C termini of BesD halogenase homologs. As above, a list of BesD halogenase homologs within a cutoff value of e^{-5} was obtained. From the list of 105 halogenases, redundant sequences of 90% identity were removed using CD-HIT, leaving 48 sequences. These remaining halogenases were input into LOGOS to generate a graphical representation of sequence conservation throughout the halogenases.

Bacterial strains. *E. coli* DH10B-T1^R was used for plasmid construction. BL21(DE3)–Star harboring the pRARE2 plasmid was used for heterologous protein production of all halogenases. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was obtained from the American Tissue Type Collection. The following strains were purchased from DSMZ: *Pseudomonas orientalis* (DSMZ 17489), *Pseudomonas trivialis* (DSMZ 14937), *Pseudomonas* sp. *Root562* (DSMZ 102504) and *Pseudomonas fulva* (DSMZ 17717). *Pseudomonas putida* KT2440 was a gift from the Keasling laboratory (University of California, Berkeley).

Construction of plasmids for protein expression. Gibson assembly was used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. PCR amplifications were carried out with Phusion polymerase or Q5 polymerase (New England Biolabs) using the oligonucleotides listed in Supplementary Table 5. The intermediate cloning plasmid, pET16-His-PrescissionCutSite-IMPDH, was constructed by amplification of IMPDH from *E. coli* genomic DNA followed by insertion into NdeI/BamHI-digested pET16b. The halogenases WP_014151497 (*S. cattleya* BesD), WP_057723975 (*P. orientalis* HalA), WP_057008702 (*P. trivialis* HalD), WP_056854138 (*P. sp. Root562* HalE) and WP_028687259 (*P. fulva* HalE), as well as LDC WP_001020973 (*E. coli*), aminotransferase 1104250A (IlvE, *E. coli*) and OCD WP_010954390 (*P. putida KT2440*) were amplified from genomic DNA of the corresponding organisms and inserted into NdeI/BamHI-digested pET16-His-PrescissionCutSite-IMPDH. The halogenases WP_016975823 (*P. fluorescens* HalA), WP_044582715 (*S. iranensis* HalB), WP_046063366 (*P. kilonensis* HalD), WP_047703910 (*P. mediterranea* HalH) were obtained as gBlocks and amplified with primers from Supplementary Table 1 before Gibson insertion into NdeI/BamHI-digested His-PrescissionCutSite-IMPDH. The remaining halogenases in Supplementary Table 1 and the lysine cyclodeaminase RapL were obtained as gBlocks containing the overhangs required for direct Gibson insertion into NdeI/BamHI-digested pET16-His-PrescissionCutSite-IMPDH. To clone mutants of *S. lavenduligriseus* BesD (SlBesD), SlBesD was amplified in two pieces from the pET16-His-SlBesD plasmid. The first piece was amplified with Sl-BesD-F and the reverse primers indicated in Supplementary Table 1. The second piece was amplified with the forward primers indicated in Supplementary Table 1 and either Sl-BesD-R (for R74A, E120A, H134A and D140A) or pET16-anneal-R (for N219A, T221A, W238A and W239A). Then both pieces were inserted into NdeI/BamHI-digested pET16-His-PrescissionCutSite-IMPDH to yield the mutant constructs. pSV272.1-His_6-MBP-PfRoot562 HalE was constructed by genomic DNA amplification with the primers indicated in Supplementary Table 1 followed by insertion into SfoI/EcoRI-digested pSV272.1. IVTT-PepC was cloned by vector amplification of the DHFR plasmid from the PURExpress IVTT kit (E640S, NEB) using the PepC-F/R primer set, followed by Gibson assembly. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences or the Barker Hall Sequencing Facility at the University of California, Berkeley.

Expression of His_{10}-tagged and His_6-MBP-tagged proteins. *E. coli* BL21 Star (DE3) cells harboring the pRARE2 plasmid were transformed with the appropriate protein-expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 l) containing the appropriate antibiotics (50 μg ml^{-1} carbenicillin or kanamycin with 50 μg ml^{-1} chloramphenicol) in a 2.8-l baffled shake flask to an optical density at 600 nm (OD_{600}) = 0.05. The cultures were grown at 37 °C at 200 rpm to OD_{600} of between 0.6 and 0.8 at which point, cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.25 mM) and overnight growth at 16 °C. Cell pellets were collected by centrifugation at 9,800*g* for 7 min at 4 °C and stored at −80 °C.

Purification of His_{10}-tagged and His_6-MBP-tagged proteins. Frozen cell pellets were thawed and resuspended at 5 ml g^{-1} cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM βME and 20% (vol/vol) glycerol, pH 7.5) supplemented with EDTA-free protease inhibitor cocktail (Roche). The cell paste was homogenized and then lysed by passage through a French pressure cell (Thermo Fisher Scientific) at 9,000 psi or by sonication with a Misonix Sonicator 3000 (power = 7.5, 5 s on, 25 s off, 2 min total process time, 0.5-inch tip). The lysate was then centrifuged at 13,500*g* for 20 min at 4 °C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (wt/vol) polyethyleneimine and stirring at 4 °C for 30 min. The precipitated DNA was then removed by centrifugation at 13,500*g* for 20 min at 4 °C. The soluble lysate was incubated with Ni-NTA (0.5 ml of resin per gram of cell paste) for 45 min at 4 °C, resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM βME and 20% (vol/vol) glycerol, pH 7.5) for 15–20 column volumes. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM βME and 20% (vol/vol) glycerol, pH 7.5). Fractions containing the target protein were pooled according to absorbance at 280 nm and concentrated using an Amicon Ultra spin concentrator (10-kDa MWCO; Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (vol/vol) glycerol and 1 mM DTT, pH 7.5) using PD-10 desalting columns. Note

that for the PLP-dependent enzymes IlvE and LDC, 20 μM PLP was present in all the purification and storage buffers. Final protein concentrations before storage are reported in Supplementary Note 2. All proteins were aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

Preparation of proteins for crystallization. For the BesD halogenase from *S. cattleya*, the protein was dialyzed following elution from the Ni-NTA column (3 × 1:50 dilution) against dialysis buffer (50 mM HEPES, 100 mM NaCl and 1 mM DTT, pH 7.5) for 1.5 h to remove imidazole. After the third round of dialysis, protein was incubated with Prescission protease (1 mg protease to 50 mg protein) and dialyzed (3 × 1:50 dilution) overnight against dialysis buffer. Cleaved and dialyzed protein was passed through Ni-NTA (2 ml) to remove Prescission and the His_{10} tag. The eluent was then diluted to a final salt concentration of 20 mM NaCl using buffer A (50 mM HEPES, 20% (vol/vol) glycerol, 1 mM DTT and 1 mM EDTA, pH 7.5) and loaded onto a 5-ml HiTrap-Q column for ion exchange with a AKTA Purifier FPLC system (GE Healthcare). The protein was eluted using a gradient from 0 to 100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, 20% (vol/vol) glycerol, 1 mM DTT and 1 mM EDTA, pH 7.5) over 40 min. The protein sample was concentrated to 2 ml and loaded onto a Superdex 75 16/60-pg (GE Healthcare) column equilibrated with SEC Buffer (20 mM HEPES, 100 mM NaCl and 1 mM DTT, pH 7.5). The protein eluent was concentrated to 15 mg ml^{-1} and glycerol was added to a final concentration of 5% (vol/vol) before flash freezing in liquid nitrogen and storage at −80 °C.

Crystallization and structure determination. Crystals of BesD from *S. cattleya* were obtained by the hanging-drop, vapor-diffusion method by combining equal volumes of protein solution (10 mg ml^{-1} BesD, lysine (1.5 mM) and αKG (3 mM, pH 7)) and reservoir solution (MES (pH 6.5, 100 mM), sodium chloride (0.6 mM) and 20% (vol/vol) PEG 4000). Thick rectangular plate crystals grew in 4 d. Crystals were transferred to an Eppendorf tube containing 100 μl of seed buffer (MES (pH 6.5, 100 mM), sodium chloride (0.6 mM) and 40% (vol/vol) PEG 4000) at 4 °C and vortexed for 30 s with 10 × 1-mm diameter zirconia–silica beads to produce a micro-seed solution. The seed solution was stored at −80 °C for future use. Fe-bound BesD crystals were prepared in a Coy anaerobic chamber by micro-seeding equal volumes of protein solution (10 mg ml^{-1} BesD, lysine (1.5 mM), αKG (3 mM, pH 7)) and reservoir solution (MES (pH 6.5, 100 mM), sodium chloride (0.6 mM) and 20% (vol/vol) PEG 4000). Thick rectangular rods grew within 2 weeks. Crystals were soaked anaerobically in a solution containing $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM), αKG (1 mM), sodium ascorbate (1 mM), lysine (1 mM) and glycerol (20% (vol/vol)) for 30 min before flash freezing in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

Data were processed with XDS[54] and scaled and merged with Aimless[55] within the CCP4 suite[56]. The anomalous signal from the Fe atoms was sufficient for obtaining initial phases using the CRANK2 pipeline[57]. Bucaneer[58] was used to build an initial model of the structure, which was refined iteratively in COOT[59] and Phenix[60]. Ligands were added to the model using COOT and refined in Phenix. The overall *B* factors of the protein are higher than that of the ligands owing to partially disordered regions of the protein. However, the *B* factors of the ligands are similar to those of the surrounding protein residues in the active site. Omit maps for the ligand complex including Fe, Cl, αKG, lysine, His137 and His204 were generated using the Phenix Map function. Fe, Cl, αKG, lysine, His137 and His204 were selected for removal from refinement before calculation of the maps. No map is visible around those ligands at the -3σ contour level. The structure was analyzed in Pymol, which was also used to create figures.

General procedure for high-resolution HPLC–MS analysis of polar metabolites with hydrophilic interaction LC. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm, 2.1 × 100 mm; EMD-Millipore) using the following buffers: buffer A (90% acetonitrile, 10% water and 10 mM ammonium formate) and buffer B (90% water, 10% acetonitrile and 10 mM ammonium formate). A linear gradient from 95% to 60% buffer A over 17 min followed by a linear gradient from 60% to 33% buffer A over 8 min was then applied at a flow rate of 0.2 ml min^{-1}. Mass spectra were acquired in positive-ionization mode using an Agilent 6530 QTOF (Agilent).

In vitro screening of BesD homolog activity. Reactions (50 μl) contained the L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-valine, L-ornithine and DL-norleucine (0.5 mM each), as well as sodium αKG (5 mM), sodium ascorbate (1 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified halogenase variants (10 μM final concentration) and allowed to proceed for 1 h at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis described above. Following identification of substrates from the amino acid pool, the reactions were repeated with individual amino acids lysine, ornithine, leucine, isoleucine and norleucine tested separately (0.5 mM each). In the case of the aliphatic halogenases that chlorinate leucine, isoleucine and norleucine, the enzymes were reassayed against each of the three amino acids separately (0.5 mM each) to confirm activity, as the substrates and products of these isomers have the same exact mass.

In vitro halogenation assay of BesD mutants. Reactions (50 μl) contained L-lysine·HCl (1 mM), sodium αKG (5 mM), sodium ascorbate (1 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified halogenase mutants or wild-type BesD from *S. lavanduligriseus* (20 μM final concentration) and allowed to proceed for 20 min at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Determination of kinetic parameters for PkHalD and PfHalA on lysine and ornithine. Reactions (100 μl) contained ATP (2.5 mM), $MgCl_2$ (5 mM), phosphoenolpyruvate (1 mM), NADH (0.3 mM), lactate dehydrogenase (10 U ml^{-1}), pyruvate kinase (10 U ml^{-1}), succinyl-CoA synthetase (3.2 U ml^{-1}), coenzyme A (1 mM), sodium αKG (1 mM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (0.2 mM), sodium chloride (10 mM) and sodium ascorbate (2 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PkHalD (5 μM) or PfHalA (5 μM) in the presence of varying concentrations of L-lysine·HCl (0–16 mM) or L-ornithine·HCl (0–2 mM). Initial rates of NADH consumption were measured by monitoring absorbance at 340 nm using a SynergyMx microplate reader (BioTek) at room temperature. k_{cat} and K_M were determined by fitting to initial rate data with Origin (OriginLab) using the equation

$$v_0 = \frac{k_{cat}[S]}{K_M + [S]}$$

where v_0 is the initial rate and $[S]$ is the substrate concentration. Analogous experiments were performed to assess the activity of PrHalE on leucine, isoleucine, norleucine and lysine.

Bromination and azidation assays. Reactions (50 μl) contained L-lysine (1 mM), sodium αKG (5 mM), sodium ascorbate (5 mM) and $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (1 mM) in 50 mM sodium acetate buffer (pH 7, adjusted with acetic acid). The anion was provided as either sodium bromide (100 mM) or sodium azide (0.5 mM). Note that halogenases used for these assays were desalted into sodium acetate (25 mM, pH 7) to prevent competition from the native chloride ligand found in the enzyme storage buffer. Reactions were initiated by addition of SwHalB (30 μM final concentration) and allowed to proceed for 1 h at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis described above.

Lysine cyclodeaminase reactions. Reactions (25 μl) contained L-lysine·HCl (1 mM), sodium αKG (5 mM), sodium ascorbate (100 μM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (50 μM) and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of SwHalB (15 μM final concentration) and allowed to proceed for 2 h at room temperature to form chlorinated lysine substrates for the lysine cyclodeaminase (RapL). Cyclodeamination was initiated by addition of 1 volume (25 μl) of cyclodeaminase solution (NAD^+ (1.8 mM) and RapL (14 μM) in 50 mM HEPES buffer pH 7.5) and allowed to proceed for 2 h at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis described above.

Ornithine cyclodeaminase reactions. Reactions (25 μl) contained L-ornithine·HCl (1 mM), sodium αKG (5 mM), sodium ascorbate (100 μM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (50 μM), and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PkHalD (15 μM final concentration) and allowed to proceed for 2 h at room temperature to form chlorinated ornithine substrate for the OCD. Cyclodeamination was initiated by addition of 1 volume (25 μl) of cyclodeaminase solution (NAD^+ (1.8 mM) and OCD (20 μM) in 50 mM HEPES buffer, pH 7.5) and allowed to proceed for 2 h at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis described above.

Lysine decarboxylase reactions. Reactions (50 μl) contained L-lysine·HCl (1 mM), sodium αKG (5 mM), sodium ascorbate (200 μM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (100 μM) and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of SwHalB (30 μM final concentration) and allowed to proceed for 2 h at room temperature to form chlorinated lysine substrates for the LDC. Decarboxylation was initiated by addition of 15 μl of LDC (48 μM) and allowed to proceed for 2 h at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on

an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis described above.

Aliphatic amino acid transaminase reactions. Reactions (50 µl) contained DL-norleucine (1 mM), sodium αKG (5 mM), sodium ascorbate (200 µM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (100 µM) and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PrHalE (10 µM final concentration) and allowed to proceed for 1.5 h at room temperature to form chlorinated norleucine substrates for the aliphatic amino acid aminotransferase (IlvE). Transamination was initiated by addition of 15 µl of IlvE (370 µM) and allowed to proceed for 1 h at room temperature before quenching in 2 volumes of methanol with 1% (vol/vol) formic acid. Samples were then analyzed by LC–MS on an Agilent 1290 UPLC-6530 QTOF (negative-ion mode) using a Phenomenex Rezex-ROA organic acid H^+ column (150 × 4.6 mM) and Carbo-H^+ security guard cartridge. In the isocratic method, the sample was eluted over the course of 10 min with 0.5% (vol/vol) formic acid.

In vitro transcription and translation of a chlorinated peptide. The IVTT reactions were performed using the PURExpress Δ(aa, tRNA) kit (NEB), which separates the amino acids and tRNA from the other reaction components. Initial halogenation reactions (4.2 µl) contained the 20 canonical amino acids (1 mM), sodium αKG (1 mM), sodium ascorbate (167 µM), $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ (167 µM) and 0.5 mM NaCl in 33 mM HEPES buffer, pH 7.5. Halogenation was initiated by addition of PfHalA or SwHalB (83 µM final concentration) and allowed to proceed for 40 min at room temperature. IVTT was then initiated by addition of the remaining PURExpress kit components along with the plasmid encoding PepC (7 ng µl^{-1} final concentration), resulting in an approximately threefold dilution of the initial halogenation reaction components. The halogenases were omitted for control reactions as noted in the data. The final reactions (14 µl) were moved to 37 °C for 4 h. Following dilution with 25 µl of 10 mM magnesium acetate, protein components were removed from the reaction by passage through a Pall Nanosep 10-kDa spin column. The flow-through was analyzed using an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 µm, 2.1 × 50 mm; Agilent) using a linear gradient from 0 to 100% acetonitrile over 5 min at a flow rate of 0.6 ml min^{-1} with 0.1% (vol/vol) formic acid as the mobile phase. Mass spectra were acquired using an Agilent 6530 QTOF.

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

Data availability

Accession codes for proteins in this study are provided in Supplementary Table 2. The PDB accession code for the BesD structure is 6NIE. Source data are available online for Figs. 3–5. Datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

49. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
50. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
51. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).
52. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).
53. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
54. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
55. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
56. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
57. Skubák, P. & Pannu, N. S. Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**, 2777 (2013).
58. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
60. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).

nature research

Corresponding author(s): Michelle C. Y. Chang

Last updated by author(s): Jul 4, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Mass Hunter Data Acquisition (B0.06.01) was used for acquisition of mass spectra on an Agilent 6530 QTOF. NMR spectra were acquired on a Bruker 900 MHz instrument with accompanying software. X-ray diffractin data were collected at Beamline 8.3.1 using a DECTRIS PILATUS3 S 6M detector and accompanying software.
Data analysis	Origin (6.0) was used for non-linear curve-fitting. Mass Hunter Qualitative Analysis (B.07.00, Agilent) was used for integration of chromatograms. NMR spectra were analyzed using Mnova (Mestrelab). X-ray diffraction data were processed with Pointless and Aimless within CCP4 (7.0.060). Data were phased with the CRANK2 software within CCP4 (7.0.060) and refined in Phenix (1.13) and COOT (0.8.9.4). Structures were visualized in Pymol (2.2.3), which was also used to make figures. The DALI server(http://ekhidna2.biocenter.helsinki.fi/dali/) was used to calculate RMSD values between BesD, SyrB2, and WelO5. MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) was used to align sequences, CD-HIT (http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi) was used to remove redundant sequences, and AliView (1.20) was used to visualizing alignments. LOGOS ((http://weblogo.threeplusone.com/create.cgi)) was used to visualize sequence conservation. BioPython(1.70) was used to sort halogenases and hydroxylases. Sequence similarity networks and genome neighborhood networks were analyzed with the publicly available web tools (https://efi.igb.illinois.edu, Cytoscape (3.6.1)) as described in the methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Accession codes for genes and proteins in this study are provided in Supplementary Tables. The PDB accession code for the BesD structure is 6NIE. Source data files for figures are provided. Datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical analyses were used to pre-determine samples size. For all chromatograms and mass spectra shown, data shown is representative of at least 3 biological replicates. The sample size was chosen to determine reproducibility based on experience of variability in these assays. For enzyme assays with quantitative data, 3 technical replicates were used to determine mean and standard deviation and is stated when relevant. Kinetic parameters (kcat and Km) were calculated using non-linear curve fitting.
Data exclusions	No data were excluded from the analyses.
Replication	Purification of proteins and of compounds for NMR analysis were performed once. For all data showing representative chromatograms and mass spectra, at least three biological replicates were performed. All reported attempts at replication were successful.
Randomization	Randomization was not relevant to this study because all experiments were performed in vitro.
Blinding	There was no blinding as this is not applicable for our sets of experiments. Blinding does not apply because experimental data are measurements of protein activity (LCMS, kinetics) or of structure (X-ray crystallography).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging