




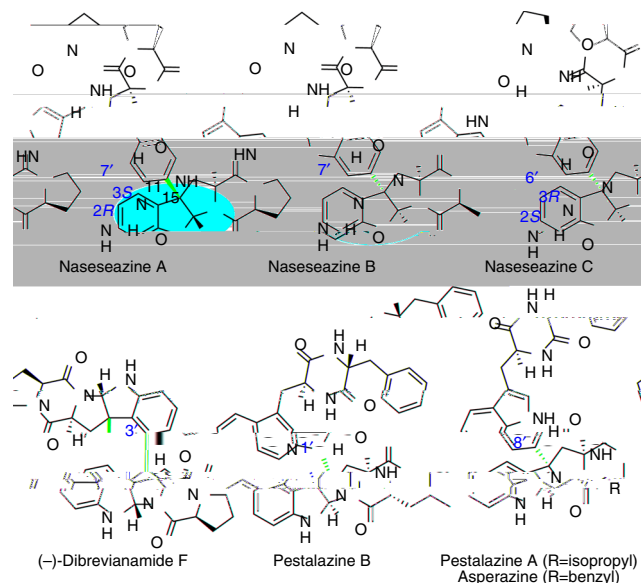
ARTICLE

# Efficient biosynthesis of heterodimeric C<sup>3</sup>-aryl pyrroloindoline alkaloids

Wenya Tian<sup>1</sup>, Chenghai Sun<sup>1</sup>, Mei Zheng<sup>1</sup>, Jeffrey R. Harmer<sup>2</sup>, Mingjia Yu<sup>3</sup>, Yanan Zhang<sup>1</sup>, Haidong Peng<sup>1</sup>, Dongqing Zhu<sup>1</sup>, Zixin Deng<sup>1</sup>, Shi-Lu Chen<sup>3</sup>, Mehdi Mobli <sup>2</sup>, Xinying Jia <sup>2</sup> & Xudong Qu <sup>1</sup>

A common heterocyclic motif observed in a large number of alkaloids and synthetic compounds is hexahydro-pyrrolo[2, 3-]indole (HPI), usually referred to as pyrroloindoline. Pyrroloindoline-containing natural products exhibit a broad array of biological properties, ranging from anticancer and antibacterial activities to the inhibition of cholinesterase<sup>1</sup>. Naturally occurring C<sup>3</sup>-aryl pyrroloindolines are mostly manifest in the fungal-sourced, tryptophan-based homodimeric diketopiperazine (DKP), in which two pyrroloindoline units are fused in a C<sup>3</sup>-C<sup>3'</sup> bond (Fig. 1)<sup>1-3</sup>. Unlike the homodimeric DKPs that are in large abundance (>100 members)<sup>1-3</sup>, only five heterodimeric DKPs have been identified so far, three of which are naseaezine A, B, and C (NAS-A, B, and C) produced by a bacterial system (Fig. 1)<sup>4-6</sup>. Except that L-Ala in NAS-A is replaced by L-Pro in NAS-B, NAS-A is identical to NAS-B. In NAS-B and C, the identical pyrroloindoline and DKP moieties are connected in two different ways: (i) the C<sup>3</sup>-aryl pyrroloindoline framework is formed through a C<sup>3</sup>-C<sup>7'</sup> bond and with 2-3 stereo-configuration (NAS-B); (ii) the connection is formed through a C<sup>3</sup>-C<sup>6'</sup> bond and with 2-3 chirality (NAS-C).

The characteristic molecular architecture and promising medicinal value of these products have garnered extensive interest particularly with respect to efforts to develop a variety of chemical methods for enantio-selective synthesis of pyrroloindoline-containing products<sup>7</sup>. However, the regio- and stereo-specificity in the densely functionalized frameworks of NASs, especially at the quaternary stereocenter at the C<sup>3</sup> position that includes an aryl substituent, requires tremendous efforts in its chemical preparation through organic synthesis<sup>8,9</sup>. This feature severely impedes an assessment of structural diversity and associated biological activities of NASs. In the chemical synthesis of NASs, regio-specificity was achieved by pre-installation of direction groups in both of the pyrroloindoline and DKP moieties, resulting in long synthetic steps and very low yields<sup>7,9</sup>. The only stereo-configuration accomplished is the C<sup>2</sup>-C<sup>3</sup>, which is induced by the C<sup>11</sup> stereocenter (derived from C $\alpha$  of tryptophan) (Fig. 1)<sup>7</sup>. However, this induction from the C<sup>11</sup> stereocenter is not able to generate the unfavored C<sup>2</sup>-C<sup>3</sup> in NAS-C. So far, no



**Fig. 1** The structures of some naturally occurring C<sup>3</sup>-aryl pyrroloindolines. Highlighted in naseaezine C is pyrroloindoline motif featuring 2S-3R chirality and C<sup>3</sup>-C<sup>6'</sup> bond connectivity. Other types of C<sup>3</sup>-aryl connectivity are highlighted with a red bond

successful strategies have been reported to chemically synthesize NAS-C. As a practical alternative to chemical synthesis of NASs, two *Streptomyces* strains have been reported to produce NAS-A, -B, and -C<sup>4-6</sup>. The biosynthesis of NAS-A, -B, and -C thus may provide an enzyme-catalyzed route for the generation of NASs through a stereo- and regio-chemically defined reaction.

Here, we unveil the biosynthesis of NAS-C and discover a key P450 enzyme, which catalyzes a highly regio- and stereo-selective C<sup>3</sup>-aryl bond-forming step to generate NAS-C. We further incorporate this P450 enzyme into a highly efficient whole-cell biocatalysis system. This engineered whole-cell factory is fed with synthetic cyclodipeptides to produce 30 heterodimeric C<sup>3</sup>-aryl pyrroloindolines (NASs analogs). Finally, some of those NASs analogs are found to have potent neuroprotective properties.

## Results

**Deciphering the biosynthesis of NASs.** Fijian marine-sourced *Streptomyces* sp. CMB-MQ030 (MQ030 strain) was previously reported to produce only NAS-A/B<sup>4,6</sup>, so our initial aim was to identify the biosynthetic gene clusters of NAS-A/B. We hypothesized that NAS-A/B were assembled by two molecules of cyclodipeptides (2,5-diketopiperazine) through a radical mechanism. In order to locate the biosynthetic gene clusters, we sequenced and assembled the draft genome of *Streptomyces* sp. CMB-MQ030 (8,454,906 bp). Analysis of the draft genome of the MQ030 strain revealed that, neighbored by several genes for regulation and exportation, each of three distinct loci (locus-1, 2, and 3) contains one tRNA-dependent cyclodipeptide synthase (CDPS) gene and one adjacent *cdps* gene, which are functionally competent for a hypothesized biosynthetic route (Fig. 2a). Locus-1 and locus-2 share high similarity to each other: 61% identity between two CDPS genes and 68% identity between the two *cdps* genes; locus-3 harbors additional albonoursin biosynthetic genes<sup>10</sup> (CDO-A and B) and is therefore excluded from further consideration.

To validate that locus-1 or -2 is the biosynthetic gene cluster of NAS-A/B, heterologous expression of locus-1 and -2 was performed in *Streptomyces* sp. J1074: no obvious metabolite was detected in the recombinant strain with locus-2, whereas the recombinant strain with locus-1 resulted in production of a single metabolite with identical molecular weight to NAS-B (Fig. 2b, trace IV). However, the high performance liquid chromatography (HPLC) retention time of this metabolite is different from that of standard NAS-B in HPLC, indicating that this metabolite is not NAS-B (Fig. 2b, trace V). This metabolite was then determined to be NAS-C by comparing the nuclear magnetic resonance (NMR) data with the reported ones<sup>5</sup>. Surprisingly, the production of NAS-C in MQ030 strain was not detected (Fig. 2b, trace III), though both NAS-C and NAS-A/B are produced by Australian marine-sourced *Streptomyces* sp. USC-636<sup>5</sup>. As locus-1 only produce NAS-C, we infer that NAS-A/B must be encoded by a different biosynthetic pathway, i.e., locus-2, though heterologous expression of locus-2 in host *Streptomyces* failed to produce NAS-A/B.

With decoupling from CDPS, the *in vitro* activity of P450 enzymes in locus-1 and -2 was also assayed to verify that the P450 works on the products of CDPS to produce NASs. It is well-characterized that CDPS catalyzes the formation of cyclic peptide dimers (cyclodipeptides), enabling us to propose that those cyclodipeptides would be substrates of the P450. Therefore, we directly feed synthetic cyclodipeptides into the P450-catalyzed reaction to assay the activity of the P450. Neither *cdps* genes from locus-1 nor -2 could be expressed in *Escherichia coli* BL21 (DE3), until their gene codons were optimized for *E. coli* usage. P450-1 (P450-NAS-C or NascB) is partially soluble in the supernatant and can be successfully purified, while P450-2 forms

inclusion bodies and was thus not characterized further. In the presence of *Fluorobacterium*-sourced flavodoxin (Fdx), flavodoxin reductase (FdxR) and an NADPH recycle system (NADP, glucose, and glucose dehydrogenase), P450-NAS-C (NascB) can convert the synthetic cyclo-(L-tryptophan-L-proline) (**s4**, cW<sub>L</sub>-P<sub>L</sub>) into NAS-C (Fig. 3a, trace III). Consistent with the *in vivo* result that no NAS-B was produced in this reaction, this enzyme is unequivocally confirmed to only be responsible for NAS-C biosynthesis.

**Delineating the mechanism of the P450 reaction.** Typical P450 reactions require the substrates to enter the active site of the P450, but not directly bind to the Fe<sup>III</sup> ion<sup>11,12</sup>. To investigate if substrate-binding causes a change in the electronic environment of the enzyme ferric heme and hence binds to the NascB, X-band continuous wave electron paramagnetic resonance (CW EPR) was used to measure the ferric heme signal. The spectra recorded at 15 K for the substrate-free enzyme (Fig. 3b, top traces) was very similar to that previously reported<sup>11</sup>, showing the enzyme predominantly in the low-spin state (LS) due to axial water coordination. The spectrum comprises a number of EPR components as indicated by the simulation (Fig. 3b, red trace) due to small differences in the orientation of the coordinated water molecule. Upon addition of cW<sub>L</sub>-P<sub>L</sub>, the majority of the ferric heme iron remains LS (Supplementary Figure 1), but only a single species is observed with shifted  $g$ -values, confirming a modification of the

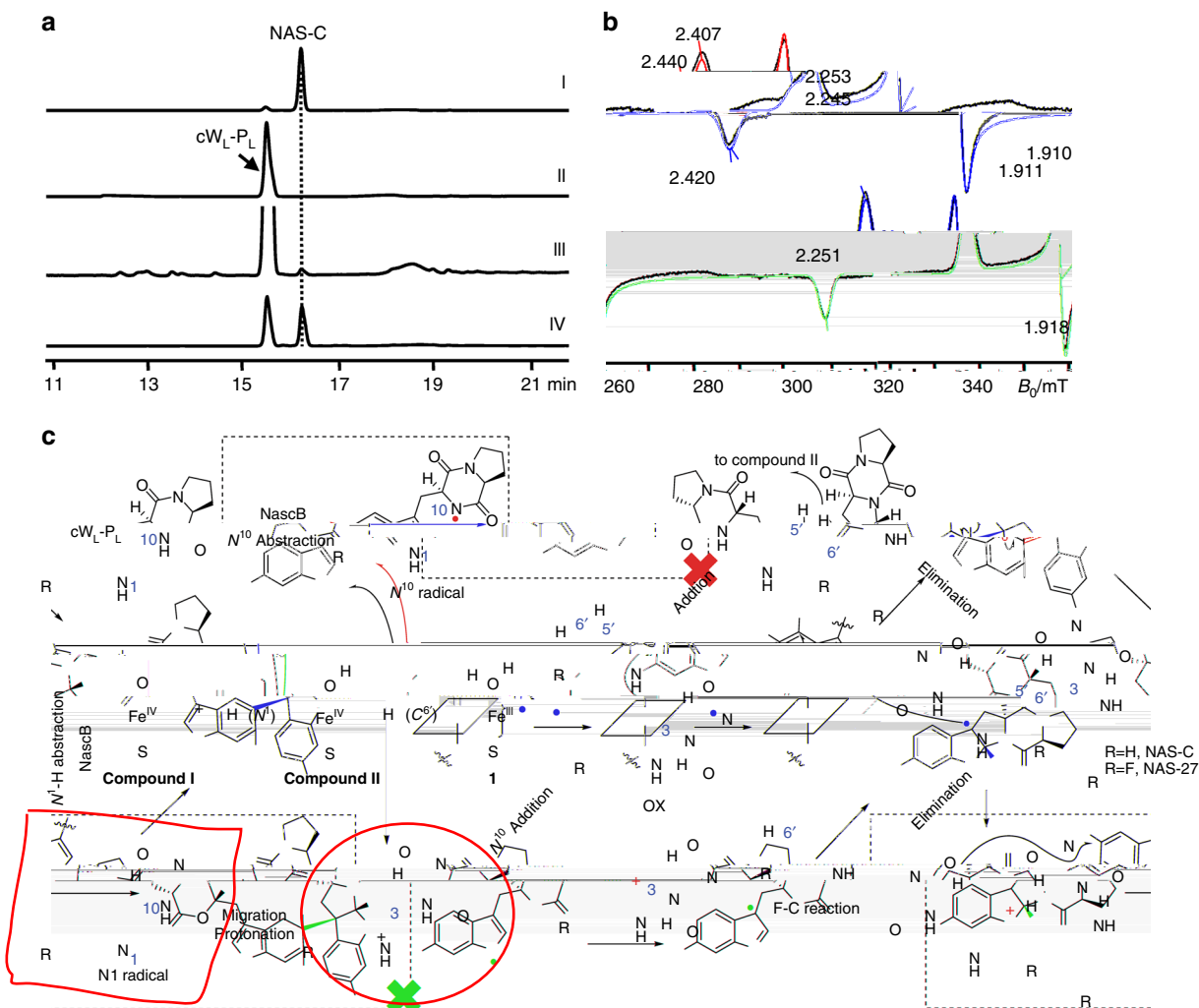
heme iron electronic environment due to the substrate cW<sub>L</sub>-P<sub>L</sub> binding to the protein.

While P450 can either catalyze oxygenation or radical-mediated coupling reactions, we hypothesize that the mechanism of forming NAS-C is a radical-mediated coupling reaction. This is supported by our result that the reaction process can be gradually inhibited by the increasing concentration of the radical scavenger TEMPO (Supplementary Figure 2). Similar to NAS-C, the biosynthesis of (-)-dibrevianamide F (homodimeric DKP) from the fungus *Flammulina velutipes* (*Fl. velutipes*) was also assumed through a radical mechanism<sup>13</sup>. Instead of CDPS, *Fl. velutipes* uses nonribosomal peptide synthase (NRPS) to synthesize the cyclodipeptide substrate and a P450 enzyme, DtpC, for catalyzing the dimerization. DtpC is proposed to initiate the reaction by abstracting a hydrogen from N<sup>10</sup> or N<sup>1'</sup> of the cW<sub>L</sub>-P<sub>L</sub> substrate to form N<sup>10</sup>• radical<sup>13</sup> or N<sup>1'</sup>• radical<sup>14</sup>. N<sup>10</sup>•-radical (N<sup>10</sup>•) undergoes intramolecular addition to C<sup>2</sup> to form the pyrroloindoline C<sup>3</sup>-radical. Two pyrroloindoline C<sup>3</sup>-radicals react with each other to yield (-)-dibrevianamide F, the major homodimeric product (Supplementary Figure 3)<sup>13</sup>. For two marginal heterodimeric heterodimeric DKPs, N<sup>1'</sup>-radical (N<sup>1'</sup>•) can either directly couple with the pyrroloindoline C<sup>3</sup>-radical to form the C<sup>3</sup>-N<sup>1'</sup> bond, or migrates to C<sup>7'</sup> first and then couple with the pyrroloindoline C<sup>3</sup>-radical to generate NAS-A/B (Supplementary Figure 3)<sup>14</sup>.

Using the density functional theory (DFT) calculations (Supplementary Methods), the structures of N<sup>10</sup>• and N<sup>1'</sup>• radical have been optimized (Supplementary Figure 4). It is revealed that the unpaired spin density in N<sup>1'</sup>• is more delocalized than in N<sup>10</sup>• (Supplementary Figure 4, Supplementary Dataset). Consequently, N<sup>1'</sup>• is more stabilized and has a free energy ( $\Delta G$ ) 16 kcal mol<sup>-1</sup> lower than N<sup>10</sup>•, implying that P450 enzymes most likely prefer the hydrogen abstraction from the cW<sub>L</sub>-P<sub>L</sub> N<sup>1'</sup> atom (instead of N<sup>10</sup>) to form the N<sup>1'</sup>•. H-atom abstraction, rather than single electron transfer, is also supported by thiolate ligation in P450 through significantly decreasing heme reduction potential and elevating pKa of compound II<sup>15,16</sup>. To further support this mechanism, we prepared an oxo-mimic of cW<sub>L</sub>-P<sub>L</sub> (Oxo-cW<sub>L</sub>-P<sub>L</sub>) (**s30** in Supplementary Figure 5), in which the HN<sup>1</sup> was replaced by an oxygen. Biochemical assay revealed that this substrate is not able to be converted by the NascB, suggesting HN<sup>1</sup> is critical for the enzymatic reaction. As assumed in the biosynthesis of communesin<sup>17</sup>, calycanthine, and chimonanthe<sup>18,19</sup>, N<sup>1'</sup>• could migrate to C<sup>3</sup>, followed by a Mannich-type reaction occurring between the N<sup>10</sup> and imine bond of N<sup>1</sup>-C<sup>2</sup>. In addition, a similar mechanism has been recently proposed in chemical oxidation of tryptophan<sup>20</sup>. Collectively, we can

the C<sup>3</sup> radical could undergo two possible routes: (1) the radical inserts into C<sup>6'</sup> of another cW<sub>L</sub>-P<sub>L</sub> followed by elimination to generate NAS-C; (2) the C<sup>3</sup> radical turns into C<sup>3</sup> cation, which attaches C<sup>6'</sup> of another cW<sub>L</sub>-P<sub>L</sub> to form NAS-C under a Friedel-Crafts scheme. We rule out the se0d route with Friedel-Crafts scheme of electrophilic aromatic substitution, because NascB-catalyzed reaction rates are not affected by strong electron-withdrawing group F on 7-F substituted cW<sub>L</sub>-P<sub>L</sub> (Fig. 3c, NAS-27 in Fig. 4, and Supplementary Figure 6).

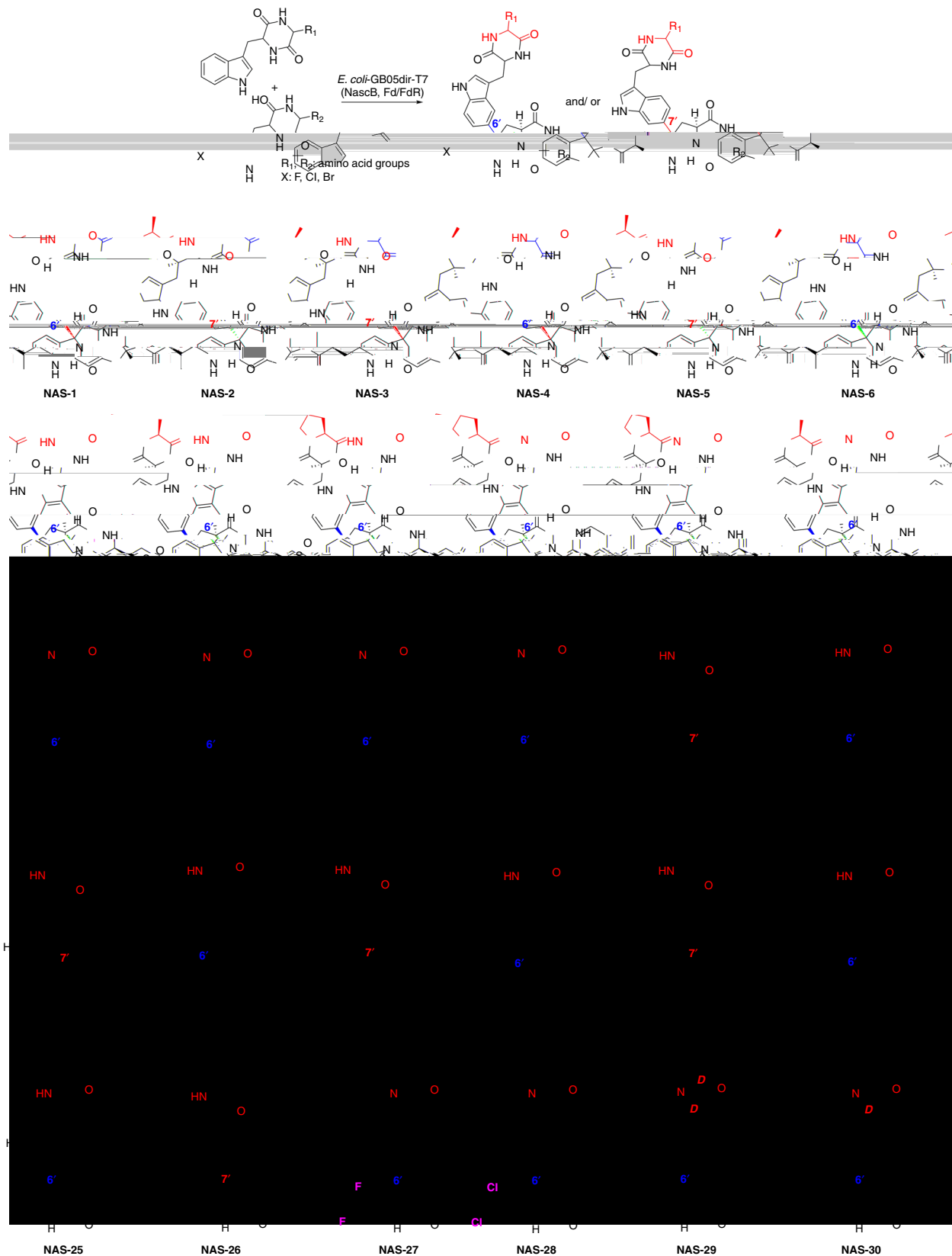
NAS-C bears a distinct bond of C<sup>3</sup>-C<sup>6'</sup>. Given that the proposed C<sup>6'</sup> radical can't be generated from the migration from N<sup>1'</sup>, the C<sup>3</sup>-C<sup>6'</sup> bond is more likely to be formed through an intermolecular (for C<sup>6'</sup>) radical additions, in which the nascent pyrroloindoline C<sup>3</sup> radical directly attacks the ( )Tj/F101Tf.51510TD(C)Tj89



**Fig. 3** Characterization of NascB in vitro and the proposed mechanism. **a** HPLC analysis of the NascB-catalyzed reaction. (I) Standard NAS-C, (II) a control reaction, the same as NascB reaction but only omitting the P450 NascB, (III) NascB reaction, using *E. coli* Fdx and FdxR as electron supply system, (IV) NascB reaction, using spinach Fd and FdR as electron supply system. **b** X-band (9.3810 GHz) CW EPR spectra recorded at 15 K showing the low-spin (LS) ferric heme signal in the absence (the upper panel) and presence (the lower panel) of the substrate (twofold excess of cWL-PL). Simulations computed using the principal  $g$ -values are shown in red: (the upper panel) two component model with  $g = (2.407, 2.253, 1.911)$  and  $g = (2.440, 2.245, 1.910)$ ; (the lower panel) single component model,  $g = (2.420, 2.251, 1.918)$ . **c** Proposed mechanisms of the radical-mediated intramolecular cyclization and intermolecular addition reaction by NascB. The NascB mechanism is proposed to involve the active species compound I (Fe<sup>IV</sup> = O+•)<sup>24–26,34</sup>. It abstracts a proton with an electron from the HN<sup>1</sup> of the substrate to form the active species compound II (Fe<sup>IV</sup>(-OH)), which is further converted into the ferric species (1) by accepting a hydrogen radical from C<sup>6'</sup> position of the substrate to form one water molecule. The water coordination turns NascB into the resting state P450, completing the catalytic cycle<sup>12,35</sup>. A significant high Gibbs free energy ( $\Delta G$ ) calculated by DFT calculation rule out the possibility of N<sup>10</sup> radical (marked by a dashed rectangle with a red cross). The possibility of Friedel-Crafts reaction from pyrroloindoline C<sup>3</sup>-radical was ruled out by the insensitivity of NascB-catalyzed reaction to the electron-withdrawing group introduced at C<sup>7</sup>-position of the substrates

this mechanism, we synthesized and assayed a variety of cyclodipeptide substrates to seek the co-generation of products with different regio-selectivity. NascB transformed the substrate cyclo-(L-tryptophan-L-valine) (cWL-VL) (s3, Supplementary Figure 5) into both products with C<sup>3</sup>-C<sup>6'</sup> (NAS-18) and C<sup>3</sup>-C<sup>7'</sup> (NAS-17) bond, respectively. Similar outcomes were also observed in NAS-4/3, 6/5, 20/19, 22/21 and 24/23 (Fig. 4). The variation of observed bonds cannot be generated by radical coupling, but rather from the addition of a pyrroloindoline C<sup>3</sup> radical to either the C<sup>6'</sup>- or C<sup>7'</sup> position. Based on these results, we confirmed that the formation of NAS-C involves a mechanism of radical-mediated intramolecular cyclization and intermolecular addition (Fig. 3c), which is very efficient for the synthesis of complex natural products.

**Developing an *E. coli*-based whole-cell biocatalysis system.** The in vitro catalyzed system requires the tedious purification of multiple enzymes, supplementation with the expensive cofactor NADPH and it is not suitable for large-scale preparation, so we decided to develop an *E. coli*-based whole-cell biocatalysis system for NAS synthesis. The NAS-C reaction depends on electron transportation systems and, therefore, we first evaluated different pairs of electron transport systems to optimize this reaction. As NAS-C can be produced by heterologous expression in *E. coli*, the endogenous ferredoxin (Fd) and ferredoxin reductase (FdxR) of *E. coli* could be competent for this reaction. Thereby, all four genes and three genes from *E. coli* were amplified and cloned into pET28a for expression in *E. coli*. Unfortunately, none of Fd and FdxR combination show activity and thus were excluded for



**Fig. 4** Products generated by the catalysis of the whole-cell system. The non-tryptophan residues in the upper moiety are highlighted in red. Connections between the upper moiety and lower pyrroloindoline moiety are indicated by two different colors: blue color for C<sup>3</sup>-C<sup>6'</sup> bond, while red color for C<sup>3</sup>-C<sup>7'</sup> bond

further study. Although we could use the *Spinacia oleracea*-sourced Fdx and FdxR, the commercial *Escherichia coli* Fd and FdR were tested and found to provide much better activity than the *Spinacia oleracea*-sourced Fdx and FdxR (~50-fold based on the conversion yield, Fig. 3a, trace IV).

Both spinach *Fdx* and *FdxR* genes were synthesized with codons optimized for *E. coli* and constructed into a variety of plasmids for evaluation of their expression yield in *E. coli*. When fused with a TRX and MBP tag at N-terminal, respectively, Fd and FdR protein can be expressed well, and the fusion proteins are fully competent without a need to remove the tags. Finally, we cloned *Fdx* and *FdxR* into a pRSFduet vector and *NascB* into pET21a to achieve the co-expression of Fd, FdR, and NascB in *E. coli*.

Unexpectedly, co-expression of these genes (*Fdx*, *FdxR*, and *NascB*) in *E. coli* BL21 (DE3) showed a highly toxic effect as cells rapidly underwent self-lysis after the expression was induced by IPTG. As cells are safe upon the individual expression of Fd, FdR, and NascB, the toxicity effect must be derived from the activated NascB in the presence of both Fd and FdR. After screening several commercial *E. coli* strains including *E. coli* Rosetta (DE3), BL21 (DE3)-pLysE, C41 (DE3), and C43 (DE3), which are widely used and some claimed to be particularly suitable for toxic protein expression, none of those strains can survive the activated NascB.

All above tested *E. coli* strains are a derivative of *E. coli* BL21 (DE3), a B type strain. The B type strains are often used for protein expression, while K type strains are mostly used for DNA cloning but also for protein expression, such as Shuffle T7 competent *E. coli* from New England Biolabs. Considering the difference between these two types, K type strain may be tolerant to the toxicity of activated NascB. We further screened two different K strains, i.e., the commercial strain *E. coli* JM109 (DE3) and a homemade strain *E. coli* GB05dir-T7. *E. coli* GB05dir is a derivative of DH10B by integration of RecET proteins on the genome<sup>21</sup>. For satisfying the requirement of expressing Fd, FdR, and NascB under T7 promoter, a T7 polymerase-coding gene was integrated into the *lacZ* operon, yielding *E. coli* GB05dir-T7. Surprisingly, the resulting strain *E. coli* GB05dir-T7 is very robust for the complete P450 catalytic system expression, while *E. coli* JM109 (DE3)-T7 died rapidly as other B type strains did. Both *E. coli* JM109 (DE3) and *E. coli* GB05dir are derived from the prototype strain K12, so their genotype is similar. The most obvious difference is the deficiency of the Lon protease in JM109 (DE3). Like other B type (DE3) strains, the *lon* gene was knocked out as it causes protein degradation. However, Lon protease is also required for stress-induced developmental changes and survival from DNA damage<sup>22,23</sup>. Because activated P450 can generate radical species, which may be able to cause DNA damage, we assume the *lon* deficiency in *E. coli* strains could be the major reason for cell death. Considering Lon can cause protein degradation, the expression yield for every single protein in GB05dir-T7 and BL21 (DE3) was compared to show the effect of Lon on protein expression is trivial: the protein yield in GB05dir-T7 is only slightly lower (~20%) than BL21 (DE3). Therefore, this strain is still efficient for protein expression and could be suitable for many other toxic P450 systems. Using this *E. coli* GB05dir-T7-based whole-cell system, complete conversion of  $cW_L-P_L$  into NAS-C can be achieved by an overnight incubation (Supplementary Figure 6a). Considering that the P450 reaction requires NADPH, we further tried co-expression of glucose dehydrogenase (GDH) with Fd, FdR, and NascB in GB05dir-T7, but the catalytic activity of NascB did not improve, suggesting that the endogenous NADPH supply in the *E. coli* is indeed sufficient.

**Generation of structural varieties through biocatalysis.** After establishing the cell biocatalysis system, we set out to perform

biocatalysis to generate NAS varieties by feeding *E. coli* cells with 20 chemically synthesized and L-Trp containing cyclodipeptide (Supplementary Figure 5), i.e.,  $cW_L-X_L$ , where  $X_L$  denotes one of 20 natural L-amino acids. To evaluate the substrate specificity of NascB, these cyclodipeptides except for the natural  $cW_L-P_L$  were individually fed to the recombinant *E. coli* (GB05dir-T7) containing the *Fdx*, *FdxR*, and *NascB* genes (GB05dir-T7-NascB). Three products were generated in high yield (Fig. 4, Supplementary Table 1, and Supplementary Figure 6b, c), including a  $cW_L-A_L$  dimerization (NAS-1) and two  $cW_L-V_L$  dimerization products (NAS-17 and NAS-18). Interestingly, NAS-1 and NAS-18 have identical connections and stereo-configuration as in NAS-C ( $C^3-C^6$  and 2-3), while NAS-17 resembles NAS-A/B ( $C^3-C^7$  and 2R-3S). The production of NAS-17 suggested that NascB indeed also has a relaxed regio- and stereo-specificity in addition to its broad spectrum of substrates.

The efficient generation of NAS-C, NAS-1, 17, and 18 suggests that substrates  $cW_L-P_L$ ,  $cW_L-A_L$ , and  $cW_L-V_L$  can be accepted readily by NascB. Considering that each of the pyrroloindolines is formed by two units of cyclodipeptides, we were interested in combining these three substrates with other cyclodipeptides to generate hetero-pyrroloindolines. Following this aim, each of these three substrates was individually co-fed with one of the remaining 17 cyclodipeptides into the recombinant *E. coli* GB05dir-T7-NascB. To our gratification, besides the production of NAS-C, -1, -17, and -18 (homo-dimerization), this co-feeding of two different cyclodipeptides at one time resulted in additional 23 products (hetero-dimerization, Fig. 4, Supplementary Tables 2–4 and Supplementary Figures 7–9): 8  $cW_L-A_L$ -derived hetero-pyrroloindolines (NAS-2 to NAS-9), 7  $cW_L-P_L$ -derived hetero-pyrroloindolines (NAS-10 to NAS-16), and 8  $cW_L-V_L$ -derived hetero-pyrroloindolines (NAS-19 to NAS-26). Interestingly, the selectivity of the upper cyclodipeptide is much stricter than the lower pyrroloindoline moiety; only  $cW_L-P_L$ ,  $cW_L-A_L$ , and  $cW_L-V_L$  can be accepted as the upper moiety. Furthermore, when co-feeding  $cW_L-P_L$  and  $cW_L-A_L$  or  $cW_L-V_L$ ,  $cW_L-P_L$  was accepted as the upper moiety (NAS-10 and 11); when co-feeding  $cW_L-A_L$  and  $cW_L-V_L$ ,  $cW_L-A_L$  was accepted as the upper moiety (NAS-2), suggesting that  $cW_L-P_L$ ,  $cW_L-A_L$ , and  $cW_L-V_L$  are the most, second, and least favored substrates, respectively, for the upper moiety. Unlike the upper moiety, the specificity for the lower pyrroloindoline moiety is more flexible and can accept in total eight tryptophan-containing cyclodipeptides, including  $cW_L-A_L$ ,  $cW_L-P_L$ ,  $cW_L-V_L$ ,  $cW_L-L_L$ ,  $cW_L-L_L$ ,  $cW_L-M_L$ ,  $cW_L-F_L$ , and  $cW_L-Y_L$ .

Besides the capability of taking various substrates, NascB also shows tolerance in the regio- and stereo-specificity of the connection manner and  $C^2-C^3$  stereo-configuration. The tolerance of these specificities is increased in the order of  $cW_L-P_L < cW_L-A_L < cW_L-V_L$  containing products: (i) In the  $cW_L-P_L$  containing products, both of the regio- and stereo-specificities are conserved and the same as those observed in NAS-C. (ii) In  $cW_L-A_L$ -containing products, six products (NAS-1, 4, and 6–9) share the same the regio- and stereo-specificity to NAS-C. Only two products NAS-2 and NAS-5 show an identical specificity to the NAS-A/B ( $C^3-C^7$  bond, 2R-3S), and one product with a not previously observed combination of bond and stereo-configuration (NAS-3,  $C^3-C^7$ , and 2-3). (iii) In  $cW_L-V_L$ -containing products, only four products (NAS-18, 20, 22, and 25) retain the specificity of NAS-C and three (NAS-17, 23, and 26) have a specificity of NAS-A/B. Three remaining products show a combination of  $C^3-C^6/2R-3S$  (NAS-24) and  $C^3-C^7/2-3$  (NAS-19 and 21), which have not previously been discovered. This substrate tolerance enables a single enzyme transformation to produce analogs with different specificities, which is very efficient for generating structural varieties. Interestingly, in some

products, the sulfur group of their methionine moiety was oxidized to a sulfoxide (**NAS-7** and **14**) or sulfone (**NAS-8**) group. This spontaneous or enzymatic oxidation (by endogenous enzymes) further increases the structural varieties of the NAS scaffold.

only trivial bioactivity has been known through testing a limited set of molecules. With our enzyme-based and efficient biocatalysis platform, it is now possible to significantly expand the bioactive space of heterodimeric C<sup>3</sup>-aryl pyrroloindolines.

## Methods

**Heterologous expression of biosynthetic gene clusters.** We extracted the genomic DNA of the NAS-producing strain of *Streptomyces* sp. (CMB-MQ030) through a method developed by Nikodinovic et al.<sup>32</sup> with minor modification. This minor modification is one more round of chloroform treatment before isopropanol precipitation of genomic DNA. To sequence the genome of the NAS-producing strain, two SMRT cells were employed at UQ Centre for Clinical Genomics (UQCCG) to generate 114,142 reads with a mean read length of 14,421 bp, which provided an average of  $\times 194.7$  coverage across the genome reference. The finished genome was assembled with HGAP2<sup>33</sup>. The gene clusters were amplified by PCR (Supplementary Methods) and inserted into the vector pB139 under the driving of the constitutive promoter (*lacI*) for heterologous expression in the model strain *Streptomyces* J1074.

**Protein expression, purification, and enzyme assay.** P450 genes with and without codon optimized for *S. aureus* were cloned into pET28a and overexpressed in *E. coli* BL21 (DE3). NascB activity against cW<sub>L</sub>-P<sub>L</sub> was assayed by incubating purified NascB (0.1  $\mu$ M) and cW<sub>L</sub>-P<sub>L</sub> (1 mM) at 18 °C with 1  $\mu$ M *S. aureus* flavodoxin (FdX), flavodoxin reductase (FdxR), or 1  $\mu$ M spinach ferredoxin (Fd), ferredoxin reductase (Fdr), 2 mM NADP<sup>+</sup> (Sigma-Aldrich), 2 mM glucose, and 2 mM glucose dehydrogenase (GDH) in 50 mM HEPES buffer, 100 mM NaCl, at pH 7.5. The reaction was left at 18 °C for 24 h. After 24 h, two times more volume of ethyl acetate was added into the reaction solution, followed by the sonication for 5 min. After the separation of aqueous and organic phase, the top ethyl acetate was transferred to a rotavapor to dry, which was re-dissolved in HPLC-graded methanol with an addition of small amount of DMSO, if the solubility is poor. The resultant solution was filtered through 0.45  $\mu$ m membrane and subjected to analysis by HPLC analysis. A diamondsil (C18, 5  $\mu$ m, 250  $\times$  4.6 mm, Dikma Technologies Inc.) was used with a flow rate at 1 mL min<sup>-1</sup> and a PDA detector over a 40 min gradient program with water (eluent A) and acetonitrile (eluent B): 0 min, 5% B; 30 min, 100% B; 33 min, 100% B; 34 min, 5% B; 40 min, 5% B.

**Electron paramagnetic resonance spectroscopy.** X-band CW EPR spectra were recorded at 15 K on a Bruker Elexsys E500 spectrometer fitted with a super high Q Bruker cavity, a liquid helium cryostat (Oxford Instrument), and a microwave frequency counter (BrukerER049X). Spectra were measured with a microwave power of 2 mW (non-saturating conditions), a modulation amplitude of 0.3 mT,



24. Belin, P. et al. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Science* **306**, 7426–7431 (2009).
25. Makino, M. et al. Crystal structures and catalytic mechanism of cytochrome P450 StaP that produces the indolocarbazole skeleton. *Science* **310**, 11591–11596 (2007).
26. Ikezawa, N., Iwasa, K. & Sato, F. Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C–C phenol coupling of (S)-reticuline in magnoflorine biosynthesis, from cultured *Aspergillus nidulans* cells. *FEBS Lett* **283**, 8810–8821 (2008).
27. Baunach, M., Ding, L., Bruhn, T., Bringmann, G. & Hertweck, C. Regiodivergent N–C and N–N aryl coupling reactions of indoloterpenes and cycloether formation mediated by a single bacterial flavoenzyme. *Chem Commun* **2013**, 9040–9043 (2013).
28. Zhang, Q. et al. Characterization of the flavoenzyme XiaK as an N-hydroxylase and implications in indolosesquiterpene diversification. *Chem Commun* **2017**, 5067–5077 (2017).
29. Präg, A. et al. Regio- and stereoselective intermolecular oxidative phenol coupling in *Streptomyces*. *Chem Commun* **2014**, 6195–6198 (2014).
30. Teufel, R., Agarwal, V. & Moore, B. S. Unusual flavoenzyme catalysis in marine bacteria. *Chem Commun* **2016**, 31–39 (2016).
31. Tang, M.-C., Zou, Y., Watanabe, K., Walsh, C. T. & Tang, Y. Oxidative