

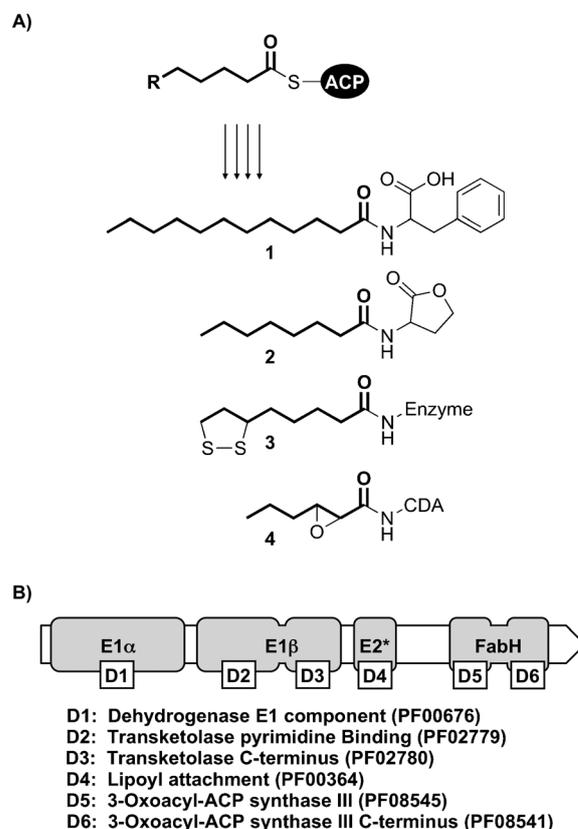
DOI: 10.1002/cbic.201100215

Discovery of a Metagenome-Derived Enzyme that Produces Branched-Chain Acyl-(Acyl-Carrier-Protein)s from Branched-Chain α -Keto Acids

Jeffrey W. Craig and Sean F. Brady*^[a]

Clones that confer the ability to produce *N*-acyl amino acids onto model cultured bacterial hosts are frequently identified in activity-based screens of soil DNA libraries. The *N*-acyl amino acids encoded by these clones are synthesized from acyl-(acyl-carrier-protein)s (acyl-ACPs) and amino acids by a diverse group of bacterial enzymes referred to as *N*-acyl amino acid synthases (NASs).^[1,2] The acyl-ACP substrates of NASs are common intermediates in the de novo synthesis of fatty acids by the type-II fatty acid synthase system (FAS II).^[3] In addition to *N*-acyl amino acids, FAS II-derived acyl-ACPs also serve as acyl donors in the production of *N*-acyl homoserine lactone autoinducers, the cofactor lipoic acid, and several lipopeptide antibiotics (Scheme 1 A, compounds 1–4).^[4–9] During a recent screen of soil metagenomic libraries for clones exhibiting antimicrobial activity, we uncovered an NAS-containing clone (EC5) that confers the ability to produce long-chain *N*-acyl-phenylalanines and *N*-acyl-tryptophans onto multiple Gram-negative host species.^[10] The NAS-encoding gene found on clone EC5, *nasA*, is the first gene in a predicted two-gene operon, *nasAB* (GenBank accession no. GQ869383). The second gene in this operon, *nasB*, encodes a putative 1118-residue protein that contains six well-characterized Pfam homology domains (Scheme 1 B).^[11] Based on generic functional predictions for these domains, we hypothesized that this enzyme was likely to be involved in the formation of ACP-linked fatty acids similar to those produced by bacterial FAS II systems. Here we report results from a series of heterologous-expression experiments in *Burkholderia graminis* C4D1M that show that NasB functions to provide branched-chain acyl-ACPs for NasA. Many bacteria, particularly Gram-negative species, do not naturally produce branched-chain fatty acids.^[3] NasB and its homologues are likely used by bacteria with straight-chain specific FAS II systems to generate branched-chain acyl-ACP substrates and should be useful as tools for synthesizing these substrates in model heterologous expression systems in which they are not natively produced (e.g., *E. coli*).

The N-terminal region of NasB contains the same set of domains as are found in the E1 α and E1 β subunits of 2-oxo acid dehydrogenase complexes (e.g., the pyruvate dehydrogenase complex, which converts pyruvate to acetyl-CoA).^[12] Following these domains is an E2-like (E2*) lipoyl attachment site. Although lipoyl attachment sites are found in the E2 subunits of 2-oxo acid dehydrogenase complexes, NasB lacks the catalytic



Scheme 1. A) Incorporation of FAS II intermediates (acyl-ACPs) into secondary metabolites and cofactors: *N*-dodecanoyl-phenylalanine (1), *N*-octanoyl homoserine lactone (2), enzyme-bound lipoic acid (3), and calcium-dependent antibiotic (CDA) (4). B) The domain architecture of NasB includes all three Pfam homology domains found in the E1 α /E1 β subunits of 2-oxo acid dehydrogenase complexes, as well as both of the Pfam homology domains found in the *E. coli* FabH protein (β -ketoacyl-ACP synthase III). Separating these two sets of domains is an E2-like (E2*) lipoyl-attachment site.

lipoamide acyltransferase domain found at the C terminus of canonical E2 subunit proteins (Pfam PF00198; exchanges the acyl-lipoate thioester with CoA, yielding a soluble acyl-CoA).^[13] Instead, the C-terminal portion of NasB contains both of the protein domains that are found in the *E. coli* FabH protein, a β -ketoacyl-ACP synthase III (KASIII)-type enzyme that performs the initial condensation reaction of FAS II (acetyl-CoA and malonyl-ACP, forming acetoacetyl-ACP).^[14]

To test our original hypothesis that NasB provides substrates for the biosynthesis of *N*-acyl amino acids, the *nasA* gene was subcloned alone and in combination with *nasB* (i.e., the *nasAB* operon) into the broad host range vector pJWC1, and these constructs were then electroporated into *B. graminis* for heterologous expression studies.^[10] Ethyl acetate extracts from cultures of *B. graminis* were examined by HPLC-MS for differences

[a] J. W. Craig, Prof. Dr. S. F. Brady
 Laboratory of Genetically Encoded Small Molecules
 The Rockefeller University
 1230 York Avenue, New York, New York 10065 (USA)
 E-mail: sbrady@rockefeller.edu

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201100215>.

in the *N*-acyl amino acids synthesized by each construct. Extracts of both cultures contained mixtures of *N*-acyl-phenylalanines, including significant amounts of the 12- and 14-carbon saturated *N*-acyl derivatives of phenylalanine (m/z 348 and 376 $[M+H]^+$, compounds **1** and **5**, respectively; Figure 1 A, B). Interestingly, the most abundant mass peak in the *nasAB* extract was from a compound with m/z 362 $[M+H]^+$ (Figure 1B), which was notably absent from the *nasA*-only extract (Figure 1A). The presence of this additional mass peak suggested that NasB afforded the production of a new 13-carbon saturated *N*-acyl-phenylalanine derivative.

To determine the nature of the presumed 13-carbon acyl chain, the metabolite corresponding to m/z 362 (**6**) was isolated from the ethyl acetate extract of *B. graminis* [*nasAB*] culture broth. Analysis of the 1D ^1H and ^{13}C NMR spectra of **6** suggested that this compound was structurally similar to previously reported *N*-acyl-phenylalanines (see the Supporting Information).^[10,15] The presence of phenylalanine was also apparent from the fragments observed in the low-resolution electrospray ionization (LRESI) MS spectrum (m/z 120 and 166). In the ^1H NMR spectrum of **6**, signals arising from the fatty acyl chain are similar to those observed in the spectra of straight-chain

N-acyl-phenylalanines isolated from *E. coli*, except that the three-proton triplet arising from the terminal fatty acid methyl group is replaced by a six-proton doublet ($\delta_{\text{H}}=0.88$ ppm, $J(\text{H,H})=6.6$ Hz; Figure 1 C, D). In the $^1\text{H},^1\text{H}$ COSY spectrum of **6**, these methyl group protons are coupled to a single methine proton (m , $\delta_{\text{H}}=1.52$ ppm) that is in turn coupled to the methylene envelope of the fatty acyl chain. Collectively, these data indicate that compound **6** is the structurally novel metabolite *N*-(11-methyldodecanoyl)phenylalanine, which contains an *iso*-branched 13-carbon saturated acyl-chain (*iso*-C13:0). The fatty acid profile of wild-type *B. graminis* does not contain branched-chain fatty acids of any type; this indicates that NasB functions to provide a source of branched-chain acyl-ACP substrates for NasA.^[16]

The production of branched-chain fatty acids has been most extensively studied in the model Gram-positive bacterium *Bacillus subtilis*, in which the FAS II system is responsible for producing both *iso*- and *anteiso*-branched fatty acids.^[17] The most important difference between the FAS II system of *B. subtilis* and that of *E. coli*, a model for straight-chain fatty acid biosynthesis, is the substrate specificity of FabH, the β -ketoacyl-ACP synthase that performs the initial condensation reaction in FAS II-mediated fatty acid biosynthesis.^[18] In *B. subtilis*, but not in *E. coli*, FabH efficiently

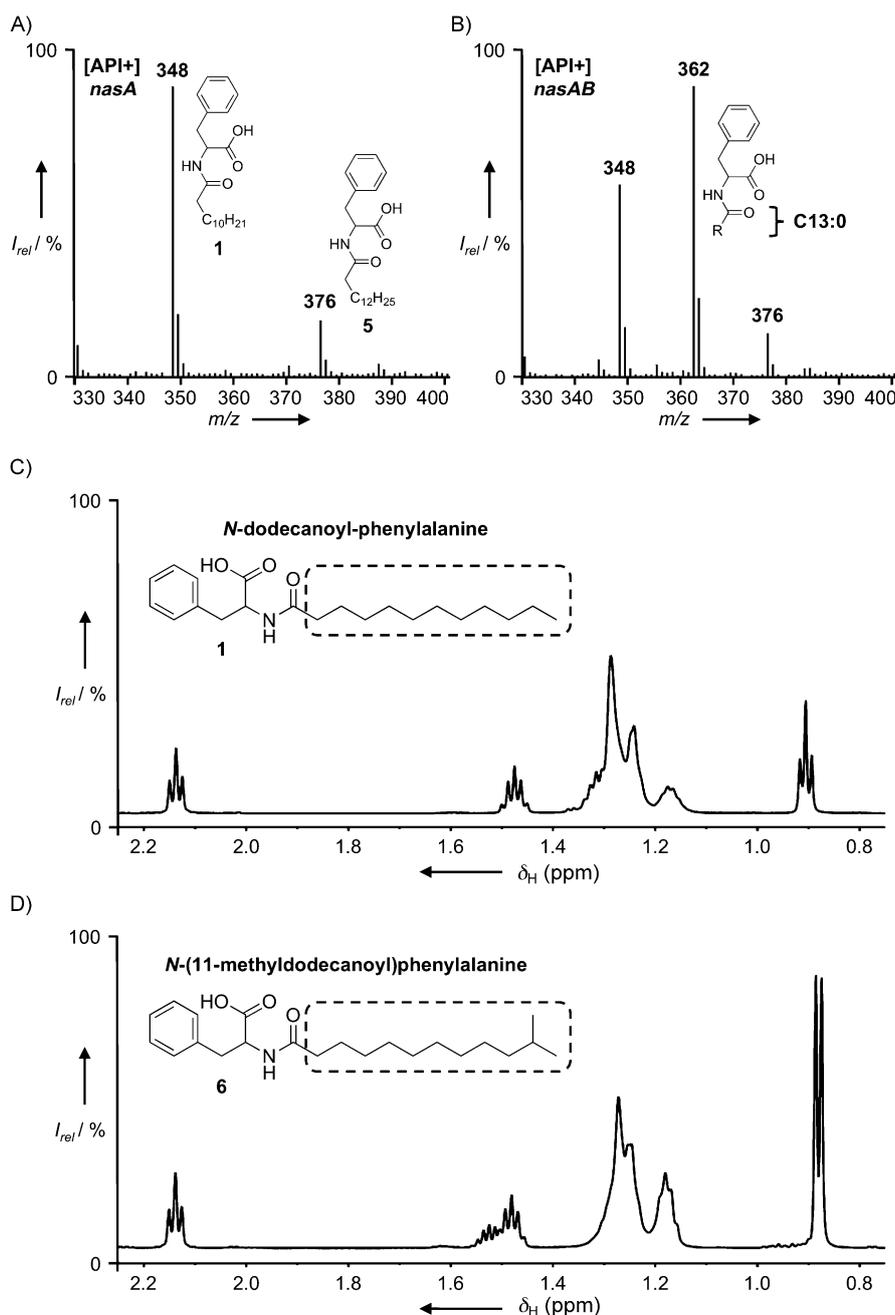


Figure 1. API-positive-mode ionization data from the HPLC-MS analysis of extracts of *B. graminis* cultures transformed with A) pJWC1-*nasA* and B) pJWC1-*nasAB*. Comparison of the fatty acid region of the 1D ^1H NMR spectrum of C) *N*-dodecanoyl-phenylalanine (**1**) with that of D) *N*-(11-methyldodecanoyl)phenylalanine (**6**).

utilizes branched-chain acyl-CoA primers. These primers are produced from α -keto acids derived from the transamination of the proteinogenic branched-chain amino acids.^[19,20] The enzymatic machinery that carries out this function is a specialized α -keto acid decarboxylase referred to as the branched-chain α -keto acid dehydrogenase complex (BCKAD).^[21] BCKAD contains the standard repertoire of protein domains that are typically found in the E1 α , E1 β , E2, and E3 subunits of other 2-oxo acid dehydrogenase complexes.

The N-terminal domains of NasB are the same as those found in the E1 α /E1 β subunits of BCKAD, thus suggesting that NasB also uses α -keto acid substrates derived from the transamination of amino acids. The odd length and terminal-branching pattern of the *iso*-C13:0 acyl-chain substructure of **6** suggested that L-leucine (**7**) was the likely source of the preferred α -keto acid substrate of NasB. To test this hypothesis and evaluate the α -keto acid decarboxylase activity of NasB, we fed 1,2-¹³C₂-labeled L-leucine to a culture of *B. graminis* [*nasAB*]. The most abundant mass peak in the organic extract of this culture was *m/z* 363 [*M*+H]⁺, one mass unit greater than that of unlabeled **6**. This result is consistent with the conversion of 1,2-¹³C₂ L-leucine to 1,2-¹³C₂ 2-oxoisocaproic acid by endogenous transaminase and the subsequent decarboxylation of this substrate by NasB to form a C1 isotopically enriched isovaleryl intermediate that feeds into the biosynthesis of **6** (compare Figure 2B with A). The mass peaks corresponding to linear-chain *N*-acyl-phenylalanines **1** and **5** (*m/z* 348 and 376 [*M*+H]⁺) were unaltered. To further confirm that the branched portion of the *iso*-C13:0 acyl chain of **6** originated from the branched side chain of L-leucine, we also fed [D₇]isopropyl-labeled L-leucine to a culture of *B. graminis* [*nasAB*]. As expected, this resulted in the appearance of a prominent mass peak with *m/z* 369 [*M*+H]⁺, which corresponded to the incorporation of the leucine-derived [D₇]isopropyl label into **6** (Figure 2C). Again, the mass peaks corresponding to linear-chain *N*-acyl-phenylalanines were unaltered. In *B. graminis*, branched-chain amino acid aminotransferase (BCAT; GenBank Accession No. EDT09763) likely performs the initial transamination reaction that provides the α -keto acid **8** used by NasB.

These feeding experiments support a model in which NasB preferentially converts the product of leucine transamination, 2-oxoisocaproic acid (**8**), to 5-methyl-3-oxohexanoyl-ACP (**10**) through an integrated series of biochemical reactions (Scheme 2A). In this model, the E1 α /E1 β -like components of NasB first decarboxylate 2-oxoisocaproic acid (**8**) to form a covalent thiamine pyrophosphate (ThPP) adduct that is subsequently transferred to the bound lipoic acid cofactor. As previously noted, NasB lacks the catalytic E2 lipoamide acyltransferase domain that would normally transfer this covalently bound intermediate from the lipoic acid cofactor to CoA. The absence of this catalytic domain suggests that instead the FabH-like component of NasB condenses malonyl-ACP (**9**) directly with the enzyme-bound isovaleryl-lipoate thioester to form 5-methyl-3-oxohexanoyl-ACP (**10**). This FAS II intermediate can then undergo three rounds of canonical fatty acid chain elongation to yield 11-methyldodecanoyl-ACP (**11**), which along

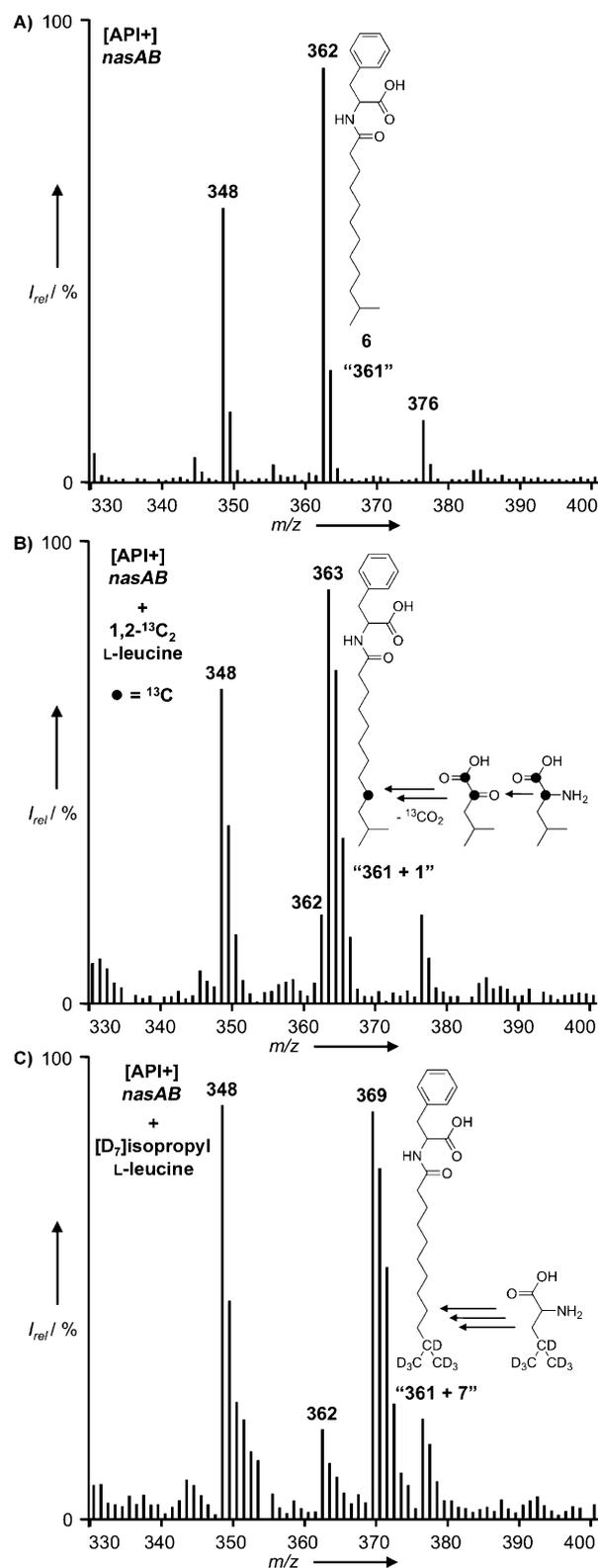
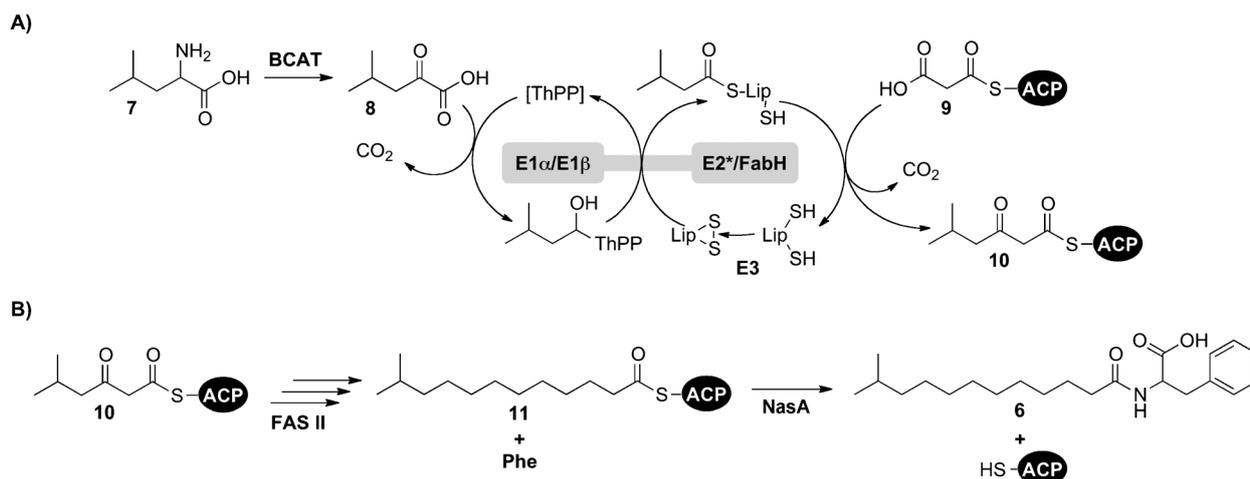


Figure 2. API-positive-mode ionization data from the HPLC-MS analysis of extracts of *B. graminis* [*nasAB*] cultures fed with either A) nothing, B) 1,2-¹³C₂-labeled L-leucine, or C) [D₇]isopropyl-labeled L-leucine.

with L-phenylalanine, is used by NasA for the production of *N*-(11-methyldodecanoyl)phenylalanine (**6**; Scheme 2B).



Scheme 2. A) The proposed condensation of 2-oxoisocaproic acid (8) and malonyl-ACP (9) by NasB to form 5-methyl-3-oxohexanoyl-ACP (10). B) Biosynthesis of *N*-(11-methyldodecanoyl)phenylalanine (6), beginning with the FAS II-mediated chain elongation of 10, yielding 11-methyldodecanoyl-ACP (11).

The *nasAB* operon structure found on clone EC5 suggests that the branched-chain acyl-ACPs produced by NasB are specifically intended for use by NasA. Branched acyl chains might therefore represent a characteristic feature used to distinguish the *N*-acyl amino acids produced by the uncultured EC5 bacterium from the more common straight-chain varieties isolated from numerous other environmental DNA clones. While the precise role(s) of *N*-acyl amino acids remain(s) to be determined, it has been proposed that they might function as signaling molecules in a manner similar to several other *N*-acylated small molecules, most notably the *N*-acyl homoserine lactones (e.g., compound 2).^[1] Variations in the acyl-chain length, degree of unsaturation, and C-3 oxidation state of *N*-acyl homoserine lactones are the primary structural determinants used to target specific populations of receptor proteins.^[22,23] Although less common by comparison, the incorporation of branched acyl chains into *N*-acyl homoserine lactones has also been reported.^[24] The above examples highlight how acyl-chain-diversification strategies can allow individual bacteria to increase the number of structurally and functionally unique metabolites they produce.

The genomes of three different species of bacteria are predicted to encode homologues of NasB containing the same putative domain architecture. These *nasB* homologues do not form operon structures with *nasA* homologues, but are instead found in divergent genetic contexts (Figure 3). The *nasB* homologue from "*Cand. A. phosphatis*" is adjacent to a homologue of the *N*-acyltransferase *lpxD*, which transfers the acyl group from 3-hydroxytetradecanoyl-ACP to UDP-3-*O*-(3-hydroxytetradecanoyl)- β -glucosamine during lipopolysaccharide biosynthesis (GenBank Accession No. ACV34570; Pfam PF04613). The *nasB* homologue from *Agrobacterium radiobacter* is adjacent to a putative fatty acid hydroxylase superfamily enzyme predicted to be involved in the oxidative modification of lipids and sterols (GenBank Accession No. ACM27362; Pfam PF04116). Although the biosynthetic abilities of the known *NasB* homologues have not been investigated, the diversity of genes to

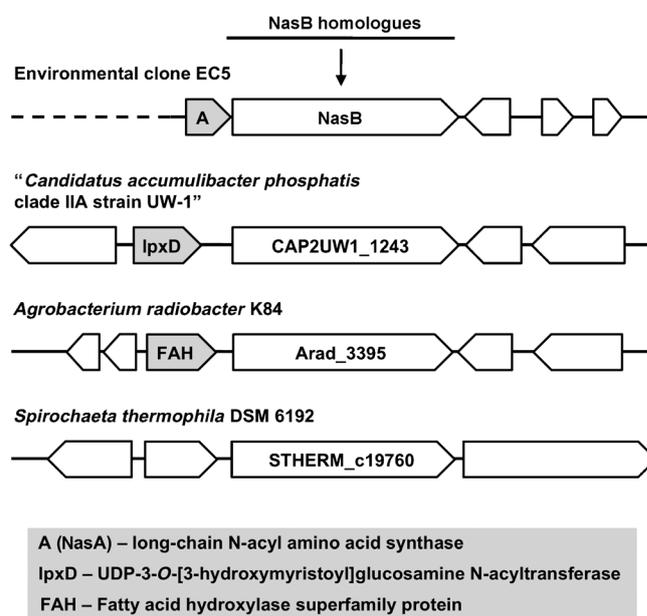


Figure 3. Putative *NasB* homologues from sequenced bacterial genomes are found in a variety of genetic contexts. The labels for *NasB* homologues are taken from the corresponding KEGG database entries. The GenBank accession numbers for CAP2UW1_1243, Arad_3395 and STHERM_c19760 are ACV34571, ACM27363, and ADN02911, respectively.

which they are linked suggests that *NasB* homologues are involved in multiple different metabolic pathways.

The aggregation of distinct enzymatic domains enhances the catalysis of multistep reactions.^[25] This is particularly true for multienzyme complexes that utilize covalent transfer as a means of channeling substrates through successive active sites and for selecting against competing metabolic pathways. In *B. subtilis* and other bacteria that produce branched-chain fatty acids, BCKAD converts α -keto acids to the corresponding acyl-CoAs, which then become substrates for *FabH*. As a covalent alternative to the CoA-based strategy used by most bacteria,

the FabH-like component of NasB might instead utilize the bound isovaleryl-lipoate thioester, thereby eliminating the need for diffusible acyl-CoA intermediates along with the potential for these intermediates to be consumed by competing catabolic pathways. Although speculative, this proposed mechanism is supported by functional inferences for each of the Pfam homology domains found in NasB, as well as by the absence of the catalytic E2 lipoamide acyltransferase domain.

To function in a heterologous host like *B. graminis*, NasB must recognize host-derived malonyl-ACP. The use of host ACP allows the acyl-ACP products of NasB to be treated as native intermediates by the endogenous FAS II system, thereby permitting subsequent utilization by acyl-ACP-requiring enzymes that fail to discriminate between branched-chain and straight-chain substrates.^[18] The introduction of *nasB* into bacterial strains that normally produce only straight-chain fatty acids could therefore provide access to branched-chain variants of numerous membrane lipids and acylated small molecules. Previous attempts to engineer *E. coli* for the production of branched-chain fatty acids have relied on the replacement of *E. coli* FabH with alternative FabH homologues derived from bacterial species that produce primarily branched-chain fatty acids. Such efforts have met with limited success.^[26] The potential for NasB to be used as a tool for the engineering of branched-chain metabolites is intriguing and warrants additional study.

Experimental Section

Bacterial culture conditions: Cultures of *Burkholderia graminis* C4D1M were grown aerobically in YENB medium (7.5 g L⁻¹ yeast extract, 8.0 g L⁻¹ nutrient broth) at 30 °C on a rotary shaker (200 rpm). Broad host-range plasmids were selected for with tetracycline (30 µg mL⁻¹). Electrocompetent *B. graminis* cells were prepared according to Sharma and Schimke.^[27]

Analytical and preparative HPLC-MS: Organic extracts were subjected to analytical HPLC-MS under the following conditions: (Waters XBridge C₁₈ 5 mm column, 4.6 × 150 mm, flow rate 1.5 mL min⁻¹); 3 min at H₂O/methanol with 0.1% formic acid (50:50), followed by a linear gradient to 100% methanol + formic acid over 12 min, followed by 100% methanol + formic acid for 5 min. Preparative HPLC-MS was performed under the same solvent conditions (Waters XBridge C₁₈ 5 mm column, 10 × 150 mm, flow rate 7 mL min⁻¹). Under these conditions, compound **6** was eluted between minutes 14.4–14.6 at a final yield of approximately 1.5 mg per L of culture broth.

N-(11-Methyldodecanoyl)-phenylalanine (6): White crystalline powder; ¹H NMR (600 MHz, [D₄]methanol, 25 °C, TMS): δ = 7.18–7.27 (m, 5H), 4.67 (dd, J(H,H) = 4.8, 9.3 Hz, 1H), 3.22 (dd, J(H,H) = 14.0, 4.7 Hz, 1H), 2.93 (dd, J(H,H) = 13.9, 9.6 Hz, 1H), 2.14 (t, J(H,H) = 7.4 Hz, 2H), 1.52 (m, 1H), 1.48 (p, J(H,H) = 7.3 Hz, 2H),

1.16–1.27 (m, 14H), 0.88 ppm (d, J(H,H) = 6.6 Hz, 6H); ¹³C NMR (150 MHz, [D₄]methanol, 25 °C, TMS) δ = 176.3, 175.1, 138.8, 130.4, 129.5, 127.9, 55.1, 40.4, 38.6, 37.0, 31.2, 30.9, 30.7, 30.6, 30.3, 29.3, 28.7, 27.0, 23.2 ppm.

Acknowledgements

This work was supported by NIH grant GM077516 and NIH MSTP grant GM07739 (J.W.C.). S.F.B. is a Howard Hughes Medical Institute Early Career Scientist.

Keywords: biosynthesis · fatty acids · metagenomics · N-acyl amino acids · natural products

- [1] S. F. Brady, C. J. Chao, J. Clardy, *Appl. Environ. Microbiol.* **2004**, *70*, 6865.
- [2] R. M. Van Wagoner, J. Clardy, *Structure* **2006**, *14*, 1425.
- [3] Y. J. Lu, Y. M. Zhang, C. O. Rock, *Biochem. Cell Biol.* **2004**, *82*, 145.
- [4] Y. H. Chooi, Y. Tang, *Chem. Biol.* **2010**, *17*, 791.
- [5] S. W. Jordan, J. E. Cronan, Jr., *J. Biol. Chem.* **1997**, *272*, 17903.
- [6] J. R. Miller, R. W. Busby, S. W. Jordan, J. Cheek, T. F. Henshaw, G. W. Ashley, J. B. Broderick, J. E. Cronan, Jr., M. A. Marletta, *Biochemistry* **2000**, *39*, 15166.
- [7] M. I. More, L. D. Finger, J. L. Stryker, C. Fuqua, A. Eberhard, S. C. Winans, *Science* **1996**, *272*, 1655.
- [8] A. L. Schaefer, D. L. Val, B. L. Hanzelka, J. E. Cronan, Jr., E. P. Greenberg, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9505.
- [9] D. L. Val, J. E. Cronan, Jr., *J. Bacteriol.* **1998**, *180*, 2644.
- [10] J. W. Craig, F.-Y. Chang, J. H. Kim, S. C. Obiajulu, S. F. Brady, *Appl. Environ. Microbiol.* **2010**, *76*, 1633.
- [11] R. D. Finn, J. Mistry, J. Tate, P. Coghill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, *Nucleic Acids Res.* **2010**, *38*, D211.
- [12] R. N. Perham, *Biochemistry* **1991**, *30*, 8501.
- [13] A. Mattevi, G. Obmolova, K. H. Kalk, A. Teplyakov, W. G. Hol, *Biochemistry* **1993**, *32*, 3887.
- [14] C. O. Rock, J. E. Cronan, Jr., *Biochemistry of Lipids and Membranes*, Benjamin/Cummings, Menlo Park, **1985**.
- [15] J. Clardy, S. F. Brady, *J. Bacteriol.* **2007**, *189*, 6487.
- [16] V. Viillard, I. Poirier, B. Cournoyer, J. Haurat, S. Wiebkin, K. Ophel-Keller, J. Balandreau, *Int. J. Syst. Bacteriol.* **1998**, *48*, 549.
- [17] T. Kaneda, *Microbiol. Rev.* **1991**, *55*, 288.
- [18] K. H. Choi, R. J. Heath, C. O. Rock, *J. Bacteriol.* **2000**, *182*, 365.
- [19] T. Kaneda, *J. Biol. Chem.* **1963**, *238*, 1229.
- [20] K. Willecke, A. B. Pardee, *J. Biol. Chem.* **1971**, *246*, 5264.
- [21] H. Oku, T. Kaneda, *J. Biol. Chem.* **1988**, *263*, 18386.
- [22] J. S. Dickschat, *Nat. Prod. Rep.* **2010**, *27*, 343.
- [23] M. K. Winson, M. Camara, A. Latifi, M. Foglino, S. R. Chhabra, M. Daykin, M. Bally, V. Chapon, G. P. C. Salmond, B. W. Bycroft, A. Lazdunski, G. S. A. B. Stewart, P. Williams, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9427.
- [24] V. Thiel, B. Kunze, P. Verma, I. Wagner-Dobler, S. Schulz, *ChemBioChem* **2009**, *10*, 1861.
- [25] L. J. Reed, *Acc. Chem. Res.* **1974**, *7*, 40.
- [26] N. Smirnova, K. A. Reynolds, *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 246.
- [27] R. C. Sharma, R. T. Schimke, *Biotechniques* **1996**, *20*, 42.

Received: March 30, 2011

Published online on June 28, 2011