HHS Public Access

Author manuscript

Chembiochem. Author manuscript; available in PMC 2015 April 01.

Published in final edited form as:

Chembiochem. 2014 March 21; 15(5): 665–669. doi:10.1002/cbic.201300794.

Identification and Characterization of A Welwitindolinone Alkaloid Biosynthetic Gene Cluster in Stigonematalean Cyanobacterium *Hapalosiphon welwitschii*

Dr. Matthew L. Hillwig, **Heather A. Fuhrman**, **Kuljira Ittiamornkul**, **Tyler J. Sevco**, **Daniel H. Kwak**, and **Prof. Dr. Xinyu Liu***

Department of Chemistry, University of Pittsburgh 219 Parkman Avenue, Pittsburgh, PA 15260, USA

Keywords

Biosynthesis; Alkaloid; Cyanobacterium; Gene cluster; Evolution

Welwitindolinones (Figure 1A) are a unique family of indole monoterpene alkaloids that were originally isolated from true-branching heterocystous filamentous cyanobacterium *Hapalosiphon welwitschii* with a broad range of biological activities.[1] In particular, (*N*-methyl)welwitindolinones B and C (**1**-**3**) are dual functional antitumor agents that exhibit antimitotic activity with multidrug resistance (MDR) reversal properties.[2] Except welwitindolinone A (**4**), all welwitindolinones isolated to date are 3,4-disubstituted oxindoles with a signature bicyclo[4.3.1]decane core motif,[1, 3] which has captivated considerable interest from the synthetic chemistry community worldwide that resulted in several total chemical syntheses of this family of molecules.[4]

Welwitindolinones are believed to be biosynthetically related to other hapalindole-type molecules as their native producers also generate hapalindoles and fischerindoles (Figure 1B),[1, 3] The biogenesis of **4**, a putative biosynthetic intermediate for bicyclo[4.3.1]decane-containing welwitindolinones, was initially proposed by Moore that involves a cationic cyclization of an oxidized 12-*epi*-hapalindole E derivative.[1] More recently, Baran and coworkers demonstrated that **4** could be readily obtained by a XeF_2-mediated oxidative ring contraction of 12-*epi*-fischerindole I (**7**),[4c] which is implicated as the direct biosynthetic precursor to **4**. However, no genes or proteins associated with welwitindolinone biosynthesis have been identified and characterized to date.

We recently initiated a collective effort in understanding the genetic and molecular basis for the biosynthesis of hapalindole-type natural products and disclosed the first biosynthetic gene cluster for ambiguine (*amb*) isonitriles.[5] Initial studies on the *amb* pathway allowed for the conclusive demonstration that the biosynthesis of ambiguines does not involve the previously proposed β-ocimene.[6] Identification of the *amb* pathway also sheds light on the roles of Rieske-type oxygenases in the late-stage structural diversification of ambiguines and

*xinyuliu@pitt.edu.

implies the involvement of AmbP1, an aromatic prenyltransferase in the generation of core hapalindole scaffold from geranyl pyrophosphate (GPP) and 3-((*Z*)-2′-isocyanoethenyl) indole (**10**, *vide infra*).

Distinct from ambiguines with a shared hapalindole G or U motif,[6-7] the hapalindoles and fischerindoles co-isolated with welwitindolinones possess different polycyclic ring frameworks, but share otherwise identical terpenoid backbone stereochemistry with hapalindoles G/U (**8**) (Figure 1C), except for an inverted C-12 quaternary stereocenter. This stereochemical discrepency at C-12 was suggested to be derived from a pair of diastereomeric monoterpene precursors, namely (*E*) and (*Z*)-β-ocimenes (**9**), according to the original biosynthetic proposal formulated by Moore and others (Figure 1D).[1, 4d, 6] Since the study of the *amb* pathway has ruled out the involvement of (*Z*)-**9** in the biogenesis of hapalindoles G/U, questions remain open on what is nature's underlying principle to generate these structural diversities. To this end, we resorted to the identification of a biosynthetic gene cluster for welwitindolinones and related hapalindoles/fischerindoles to gain additional insights on the biogenesis of hapalindole-type natural products.

We chose *H. welwitschii* UTEX B1830, a xenic strain to investigate the biosynthesis of welwitindolinones as it is readily available in the public domain and has been reported to produce identical hapalindole-type molecules as *H. welwitschii* W. & G .S. West.[1] We initially examined the metabolite profiles of *H. welwitschii* UTEX B1830 by combining HPLC with UV-spectral fingerprints and high resolution mass spectral (HRMS) analyses (Figure SI-1B) and ensured it generated the structural diversities as previously reported.[1] We then extracted the genomic DNA of *H. welwitschii* UTEX B1830 and subjected it to *de novo* genome sequencing using a Roche 454 GS FLX+ system (SI Methods). The draft assembly of total reads resulted in nearly 10,000 contigs that total 15 Mbp, confirming the xenic status of *H. welwitschii* UTEX B1830. Using this pseudometagenomic data, we carried out nucleotide BLAST using genes in the *amb* pathway as bioinformatic leads. This effort led to the identification of 11 contigs, including a single 21-kbp contig that resembles the genetic sequence from *amb*C1 to *amb*C3 in the *amb* gene cluster, with the remaining contigs having an average size of 1-2 kbp and lacking homologous end-joining sequences. Subsequent gap repairings relied extensively on Sanger sequencing of carefully designed cross-contig PCR amplicons (SI Methods) in order to bypass highly sequence-repetitive regions (Figure SI-2) to successfully map out the sequence and directionality of the entire welwitindolinone (*wel*) biosynthetic gene cluster that spans 36 kbp long (Figure 2).

Functional annotation of 30 protein-coding open reading frames (ORFs) in the *wel* gene cluster revealed striking similarity to those in the *amb* pathway (Figure 2 & Table SI-1), providing an initial glimpse on two highly related biosynthetic machineries for the assembly of welwitindolinones, ambiguines and related hapalindoles. The presence of transposable elements (*wel*S1/S2) in the *wel* cluster, similar to *orf*2/3 at the boundary of the *amb* cluster, highlights the mobile nature of these pathways, suggesting horizontal gene transfer (HGT) may be responsible for the wide occurrence of hapalindole-producing cyanobacteria in the Stigonematalean family. Except transposase-coding *wel*S1-2 and response regulator-coding *wel*R3, proteins encoded by 19 ORFs that span 27 kbp long from *wel*D4 to *wel*C1 share extremely high sequence identities (90-99%) to those in the *amb* gene cluster, implicating

they are likely functionally identical to their *amb* homologues in the context of regulating and assembling key biosynthetic intermediates for welwitindolinone and ambiguine biogenesis (Figure 3A). To correlate with the bioinformatic predictions, we overexpressed WelP1 and WelP2 in *E. coli*, and demonstrated that WelP2 is a dedicated geranyl diphosphate synthase to provide GPP (SI Methods), whereas both WelP1 and WelP2 lack the ability to convert GPP to β-ocimene, identical to what was observed for AmbP1 and AmbP2 in ambiguine biosynthesis.[5] To validate the functions of *wel*I1-3, we cloned them as a single operon under a pBAD promoter (SI Methods) and examined its biosynthetic product *in vivo*. Overexpression of *wel*I1-3 and *amb*I1-3 in *E. coli* both led to the robust production of **10** that matched the synthetic standard (Figure 3B) by HPLC analysis with no *E*-isomer of **10** observed. These experiments collectively demonstrated that welwitindolinone and ambiguine biosynthesis adopt identical pathways for assembling early biosynthetic intermediates GPP and **10**.

Upstream of *wel*D4 in the *wel* cluster there are eight ORFs that show more notable differences, compared to those embedded in the *amb* pathway (Figure 2). In particular, *wel*M is unique to the *wel* gene cluster and predicted to encode a SAM-dependent methyltransferase, suggesting it is likely responsible for the generation of *N*-methylwelwitindolinones (**1b**, **3b**) from their non-methylated precursors (**1a**, **3a**) (Figure 4A). To validate the function of WelM, we overexpressed and purified it from *E. coli* (SI Methods) and isolated its putative substrate **3a** (Figure SI-3) from *H. welwitschii* UTEX B1830. Incubation of **3a** with recombinant WelM and S-adenosylmethionine (SAM) rapidly generated a new product, of which the retention time over a C18 HPLC column matches that of authentic **3b** (Figure 4B). Thorough characterizations of the enzymatic product derived from **3a** and WelM by 1D/2D NMR and HRMS analysis confirmed its structural identity to be **3b** (Figures SI-4/5/6). The kinetic profiles of WelM (K_m=2.43±0.18μM and k_{cat}=2.17±0.12s^{-1} for **3a**; SI Methods) are in line with other recently characterized amide N-methyltransferases,[8] providing additional evidence that **3a** is a natural substrate for WelM. We also examined the substrate promiscuity of WelM, motivated by the observations that several methyltransferases in small molecular natural product biosynthesis possess broad substrate tolerance. [9] Among a broad range of substrates tested, including non-substituted oxindole, 3-methyl oxindole and several readily isolatable hapalindole-type molecules (Figure SI-7A), only **4** that contains an oxindole backbone appended with a spirocyclobutane monoterpene unit turned out to be an acceptable substrate for WelM (Figure SI-7B). However, **4** is a very poor substrate with an apparent k_{cat} ca. 1,000-fold less than that for **3a**. Overall, these experiments demonstrate that WelM is a highly efficient narrow-substrate N-methyltransferase that was likely evolved to specifically recognize 3,4-disubstituted oxindoles with a bicyclo[4.3.1]decane motif. This dedicated N-methylation is likely the final step for the welwitindolinone biosynthesis and allows for the generation of **3b** with reduced cytotoxicity and enhanced MDR activity in comparison with its precursor **3a**.[2a] The dedicated activity of WelM constitutes an excellent example of how nature uses late-stage tailoring enzymes to diversify its product inventory with enhanced biological profiles.

The remaining biosynthetic genes embedded in the *wel* cluster encode five nonheme iron (NHI)-dependent oxygenases, including four full length Rieske-type oxygenases (WelO1-O4) and a Fe(II)/α-ketoglutarate-dependent oxygenase (WelO5). While the number and diversity of *wel* oxygenases mirrors those in the *amb* pathway, their protein sequence identities are visibly lower (61-79%) (Table SI-1), in comparison with the rest of *wel*/*amb* biosynthetic enzymes (*vide supra*). Classification of fifteen welwitindolinones and 12-*epi*-hapalindoles/fischerindoles derived from the *wel* pathway based their oxidation states at the indole terpenoid cores (Figure 5A) clearly illustrates a need of five distinct 2e oxidation events to complete the full oxidative maturation of welwitindolinones. This analysis also re-asserts the proposed role of WelO5 as a candidate NHI-dependent halogenase for the stereoseletive generation of a C-Cl bond at the C-13 of hapalindole/fischerindole backbones as required for the generation of **5c**/**5d**/**6c** from **5a**/**5b**/**6a** respectively (Figure 5B). WelO1-4 are four highly homologous Rieske oxygenases (Figure SI-8), which are likely the main catalysts for the oxidative transformations of fischerindoles to welwitindolinones (Figure 5C), based on oxidation state analyses and a recently disclosed nonenzymatic oxidative transformation of **6c** to **4**.[4c] In addition, distinct from Rieske oxygenases in the *amb* pathway, the iron-sulfur domain sequences in WelO1-4 are virtually identical (Figures SI-2/SI-8), implicating they likely originate from a common ancestor during HGT and are evolved specifically around the NHI catalytic domains.

Functional characterizations of WelI1-3/WelP2/WelM as well as bioinformatic correlations of WelO1-5 with the structural diversities of hapalindole-type molecules derived from *H. welwitschii* UTEX B1830 provided strong biochemical supports for linking the *wel* pathway to welwitindolinone biosynthesis. Comparative analysis of the *wel* and *amb* pathways collectively demonstrates the roles of GPP and 3-((*Z*)-2′-isocyanoethenyl) indole as common intermediates for the biosynthesis of welwitindolinones and ambiguines, and the critical roles of NHI-dependent oxygenases, particularly Rieske oxygenases as the key catalysts for late-stage oxidative structural diversifications.

The *in vitro* characterization of WelI1-3/WelP2 and inability of WelP1/P2 to generate β-ocimine, as shown in the *amb* pathway, solidified our conclusion that the stereodivergent generations of 12-*epi*-hapalindoles/fischerindoles from the *wel* pathway, in contrast to ambiguines with a conserved hapalindole G/U motif from the *amb* pathway, is not derived from the stereochemical differences of biosynthetic precursors (Figure 1D), as those proposed in the literature.[1, 4d, 6]. Early studies on the *amb* pathway by us implicate a possible role of AmbP1 in directly fusing GPP and **10** to **8a** through a cationic cascade.[5] However, as the sequences of WelP1 and AmbP1 are virtually identical (Table SI-1), it seems unlikely that they will be the sole catalyst responsible for the stereodivergent generation of **5a, 6a** and **8a**. In light of the recent disclosures of oxygenase-mediated terpene-like cyclizations of geranylated aromatic polyketides,[10] it is plausible that both WelP1 and AmbP1 will mono-geranylate **10** with GPP and subsequently rely on a dual functional oxygenase in the AmbO or WelO families. In addition, the *wel* pathway does not contain a dedicated gene for transferring a nucleophilic “sulfur” to isonitrile to generate isothiocyanate group. It is possible that a pathway-independent sulfur carrier protein (such as ThiS in thiamin biosynthesis) or an endogenous cystein desulfurase harboring a reactive

thiocarboxylate or a cystein persulfide may serve as the sulfur donor,[11] as recently proposed for 2-thiosugar biosynthesis.[12] Efforts in delineating the remaining mysteries associated with hapalindole-type natural product biosynthesis are currently undergoing in our laboratory.

Experimental Section

Experimental details, including SI Methods, Figure SI-1 to SI-10, Table SI-1&2, are described in the supporting information. Nucleotide sequence of the *wel* cluster was deposited in Genbank with accession number KF811479.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported in part by University of Pittsburgh – Department of Chemistry startup fund, the Clinical and Translational Science Institute (CTSI) via NIH Grants UL1RR024153 and UL1TR000005, the Competitive Medical Research Fund of University of Pittsburgh Medical Center and University of Pittsburgh Honors College (fellowships to HAF and TJS).

References

[1]. Stratmann K, Moore RE, Bonjouklian R, Deeter JB, Patterson GML, Shaffer S, Smith CD, Smitka TA. J. Am. Chem. Soc. 1994; 116:9935.

[2]. a) Smith CD, Zilfou JT, Stratmann K, Patterson GML, Moore RE. Mol. Pharmacol. 1995; 47:241. [PubMed: 7870031] b) Zhang X, Smith CD. Mol. Pharmacol. 1996; 49:288. [PubMed: 8632761]

[3]. Jimenez JI, Huber U, Moore RE, Patterson GM. J. Nat. Prod. 1999; 62:569. [PubMed: 10217710]

[4]. a) Baran PS, Richter JM. J. Am. Chem. Soc. 2005; 127:15394. [PubMed: 16262402] b) Reisman SE, Ready JM, Hasuoka A, Smith CJ, Wood JL. J. Am. Chem. Soc. 2006; 128:1448. [PubMed: 16448105] c) Baran PS, Maimone TJ, Richter JM. Nature. 2007; 446:404. [PubMed: 17377577] d) Richter JM, Ishihara Y, Masuda T, Whitefield BW, Llamas T, Pohjakallio A, Baran PS. J. Am. Chem. Soc. 2008; 130:17938. [PubMed: 19035635] e) Bhat V, Allan KM, Rawal VH. J. Am. Chem. Soc. 2011; 133:5798. [PubMed: 21446729] f) Huters AD, Quasdorf KW, Styduhar ED, Garg NK. J. Am. Chem. Soc. 2011; 133:15797. [PubMed: 21819133] g) Quasdorf KW, Huters AD, Lodewyk MW, Tantillo DJ, Garg NK. J. Am. Chem. Soc. 2012; 134:1396. [PubMed: 22235964] h) Allan KM, Kobayashi K, Rawal VH. J. Am. Chem. Soc. 2012; 134:1392. [PubMed: 22235963] i) Styduhar ED, Huters AD, Weires NA, Garg NK. Angew. Chem. Int. Ed. 2013; 47:12422.

[5]. Hillwig ML, Zhu Q, Liu X. ACS Chem. Biol. 2014 10.1021/cb400681n.

[6]. Raveh A, Carmeli S. J. Nat. Prod. 2007; 70:196. [PubMed: 17315959]

[7]. Smitka TA, Bonjouklian R, Doolin L, Jones ND, Deeter JB, Yoshida WY, Prinsep MR, Moore RE, Patterson GML. J. Org. Chem. 1992; 57:857.

[8]. a) Wu Y, Kang Q, Shang G, Spiteller P, Carroll B, Yu T-W, Su W, Bai L, Floss HG. ChemBioChem. 2011; 12:1759. [PubMed: 21681880] b) Giessen TW, von Tesmar AM, Marahiel MA. Biochemistry. 2013; 52:4274. [PubMed: 23705796]

[9]. a) Pacholec M, Tao J, Walsh CT. Biochemistry. 2005; 44:14969. [PubMed: 16274243] b) Zhang C, Albermann C, Fu X, Peters NR, Chisholm JD, Zhang G, Gilbert EJ, Wang PG, Van Vranken DL, Thorson JS. ChemBioChem. 2006; 7:795. [PubMed: 16575939] c) Lee J-H, Bae B, Kuemin M, Circello BT, Metcalf WW, Nair SK, van der Donk WA. Proc. Natl. Acad. Sci. 2010; 107:17557. [PubMed: 20876132]

[10]. a) Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S. FEBS Lett. 2007; 581:2929. [PubMed: 17544411] b) Chooi Y-H, Hong YJ, Cacho RA, Tantillo DJ, Tang Y. J. Am. Chem. Soc. 2013; 135:16805. [PubMed: 24161266]

[11]. a) Begley TP, Xi J, Kinsland C, Taylor S, McLafferty F. Curr. Opin. Chem. Biol. 1999; 3:623. [PubMed: 10508664] c) Mueller EG. Nat. Chem. Biol. 2006; 2:185. [PubMed: 16547481]

[12]. Sasaki E, Liu H.-w. J. Am. Chem. Soc. 2010; 132:15544. [PubMed: 20961106]

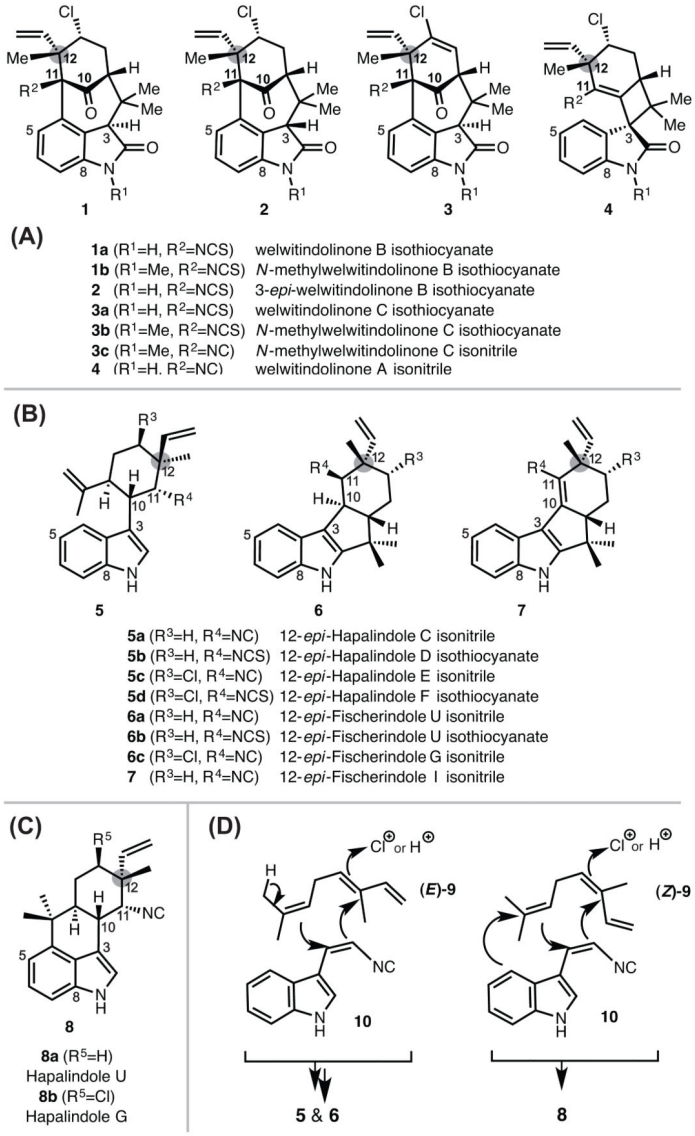

Figure 1.

Structures of welwitindolinones (A) and 12-*epi*-hapalindoles/ fischerindoles (B) isolated from *H. welwitschii* W. & G .S. West and UTEX B1830 (C-12 stereocenters are highlighted) and their distinctions from hapalindoles G/U (C), the backbone structural motifs in ambiguines (C-12 stereocenters are highlighted). (D) Biosynthetic proposals on the stereodivergent generations of **5a/c**, **6a/c** and **8** from (*E*)- or (*Z*)-**9** with **10,** formulated by Moore and others.[1, 4d, 6]

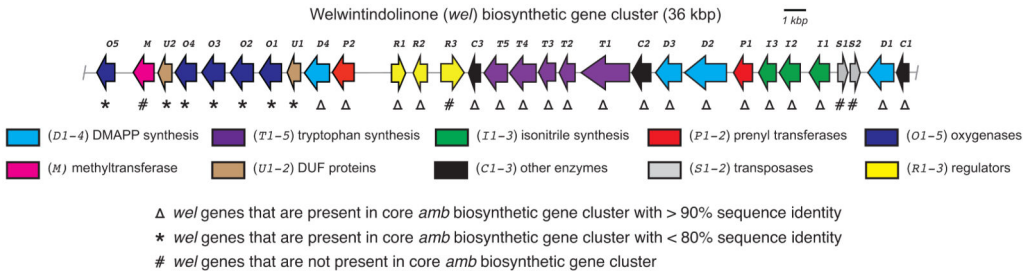

Figure 2.

Illustration of the welwitindolinone (*wel*) biosynthetic gene cluster from *H. welwitschii* UTEX B1830 and its comparison with the ambiguine (*amb*) biosynthetic gene cluster from *F. ambigua* UTEX1903. *Wel* gene functions are grouped based on their putative roles associated with welwitindolinone and 12-*epi*-hapalindole/fischerindole biosynthesis and their coded protein sequences were compared with those in the *amb* pathway for similarities (highlighted with symbols Δ/*/#). Full annotation of *wel* ORFs can be found in the supporting information (Table SI-1)

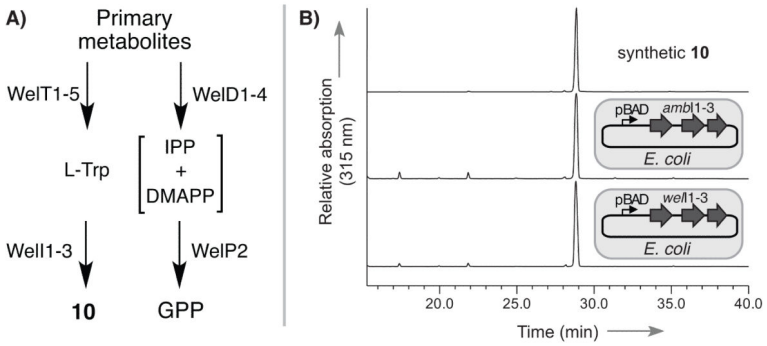

Figure 3.

Prediction and characterization of *wel* genes involved in the assembly of intermediates GPP and **10** for welwitindolinone biosynthesis. (A) Predicted functions of WelD1-4, WelT1-5, WelI1-3 and WelP2. (B) *In vivo* characterization of WelI1-3 enzymatic product in *E. coli* and its comparison with that derived from AmbI1-3.

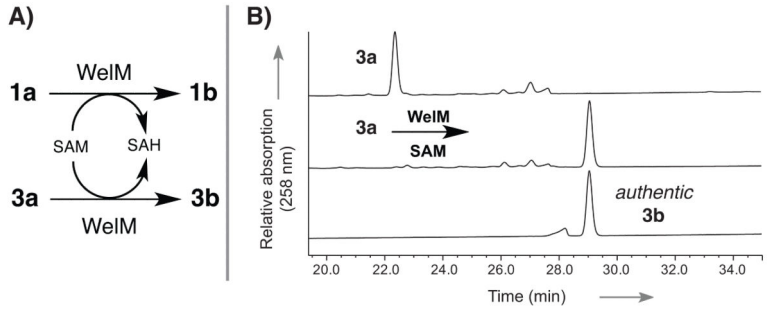

Figure 4.

Prediction and characterization of WelM involved in a late stage N-methylation for welwitindolinone biosynthesis. (A) Proposed function of WelM in the generation of N-methylated welwitindolinones; (B) Characterization of WelM and its enzymatic product derived from **3a**.

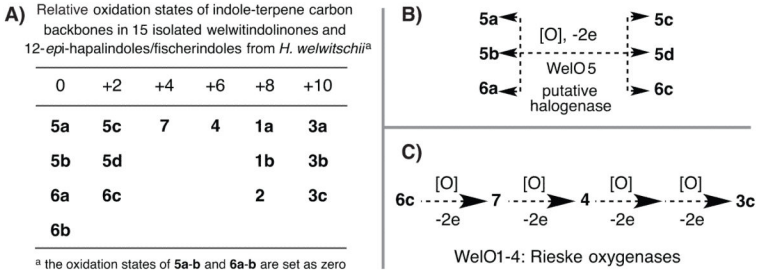

Figure 5.

(A) Classification of welwitindolinones and 12-*epi*-hapalindoles/fischerindoles isolated from *H. welwitschii* based on the oxidation state of core indole-monoterpenoid scaffold. (B) Proposed role of WelO5 as a Fe(II)/α-ketoglutarate dependent halogenase in the generation of **5c**/**5d**/**6c** from **5a**/**5b**/**6a**. (C) Proposed roles of WelO1-4 (Rieske-type oxygenases) in the oxidative maturation of welwitindolinones from 12-*epi*-fischerindoles. A putative pathway to generate **3c** from **6c** by four sequential 2e-oxidations was shown.