Electronic Supplementary Material (ESI) for Chemical Communications.

Electronic Supporting Information

Molecular and genetic basis for early stage structural diversifications in hapalindole-type alkaloid biogenesis

Qin Zhu[a] and Xinyu Liu[a]*

[a]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260

*To whom correspondence should be addressed: xinyuliu@pitt.edu

ESI Materials and Methods

General methods. All polymerase chain reactions (PCRs) were carried out on a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by Elim BioPharm Inc. Preparative-scale reverse-phase HPLC was performed using a Dionex instrument equipped with Luna C18 columns (21 x 250 mm and 4.6 x 250 mm) (Phenomenex). Analytical reverse-phase HPLC was performed using a Dionex UHPLC with a photo-diode array UV/Vis detector (Thermo Fisher Scientific) and a 4.6 x 250 mm Luna C18 column (Phenomenex). HRMS analysis was conducted using a Q Exactive Benchtop Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a Dionex RSLC (Thermo Fisher Scientific). NMR spectrum was recorded on a Bruker Avance III 700 MHz spectrometer equipped with a $^{1}H/^{13}C/^{15}N$ triple-resonance inverse probe (1.7mm 'microprobe').

Materials. Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kappa HiFi DNA polymerase was obtained from Kappa Biosystems. Restriction endonucleases, T4 DNA ligase and Antarctic phosphatase were purchased from New England Biolabs. LB broth and agar used for culturing *E. coli* were obtained from Teknova. All other reagents including inorganic salts and cofactors were purchased from Sigma-Aldrich or Fisher Scientific unless otherwise stated.

Strains and plasmids. *E. coli* TOP10 cell (Life Technologies) was used for routine cloning and plasmid propagation. *E. coli* C43(DE3) cell (Lucigen) was used for protein expression. pQTEV cloning plasmid was obtained from Addgene.

GenBank# for WelU1, WelU2, WelU3, AmbU4, WelC1 and WelC3 proteins described in this work: **WelU1:** AHI58823 **WelU2:** AHI58818 **WelU3:** AIH14742 **AmbU4:** AHB62757

WelC1: AHI58845 **WelC3:** AHI58829

Protein expression. Genes coding WelU1, WelU2, WelU1(27-228), WelU2(27-226), WelC1 and WelC3 proteins were amplified from *H. welwitschii* UTEX B1830 genomic DNA using oligo pairs 1/2, 3/4, 5/2, 6/4, 7/8 and 9/10 (Table S1) respectively. They were ligated to the BamHI/NotI sites of cloning vector pQTEV. Positive clones were identified by restrictive digestion and sequence verified. WelU3-coding gene was synthesized by Bio Basic Inc (Ontario, Canada) and ligated to pQTEV analogously. Each pQTEV construct was transformed to *E. coli* C43(DE3) cells (Lucigen). Subsequent heterologous expressions and purifications were carried out in an identical fashion as described for AmbP1 and WelP1 proteins in reference 1 using IMAC. The purified proteins were analyzed by SDS-PAGE to ensure homogeneity (Figure S10), concentrated with 10 kDa cutoff concentrator tubes, assayed, and the remainder was flash-frozen using liquid nitrogen and stored at -80 C for later assays. Approximate yields for each protein are: WelU1 (1.0 mg/L), WelU2 (insoluble), WelC1 (1.2 mg/L), WelC2 (1.1 mg/L), WelU1(27-228) (5.0 mg/L), WelU2(27-226) (3.8 mg/L), WelU3 (1.5 mg/L).

Initial in vitro assay with WelC1, WelC3, WelU1 using substrate 1: To examine whether WelC1, WelC3, WelU1 has the ability to convert **1** to **2**. Assays in the scale of 100 µL were set up at pH=6.5 (MES buffer) and 7.5 (HEPES buffer) with 0.5 mM of **1** in the presence of either WelC1, WelC3, or WelU1 (10 µM, final concentration). The enzymatic reactions were allowed to proceed at 30 ºC for 6 hour before being quenched with EtOAc extraction. EtOAc layer was dried under a stream of nitrogen gas, redissovled in MeOH and subjected to HPLC analysis.

In vitro assays with WelU1, WelU2, WelU3, WelU(27-228) and AmbU4: For a typical standalone assay, substrate **4** was synthesized using WelP1 and immediately used for the subsequent reactions. A100 µL reaction was set up that contains 1 mM **3**, 1.5 mM GPP, 10 uM WelP1 and 20 mM $MgCl_2$ in 50 mM Tris buffer (pH 9.0). The reaction was incubated at 30 °C for 1h and quenched by a 1 mL

EtOAc extraction. The EtOAc layer was dried under a stream of nitrogen. To this crude preparation of **4** was added the buffer at the desired pH (MES buffer for pH=5.0, 6.0 and Tris buffer for pH=7.0, 8.0 and 9.0) with 10% DMSO to 100 µL and the U enzyme (5 µM final concentration). The reaction was incubated at 37 °C for 5-120 min and quenched with 1 mL EtOAc extraction. The EtOAc fraction was dried and dissolved in 100 uL MeOH for HPLC analysis. For a coupled assay with WelP1, a 100-µL reaction was set up at either pH=9.0 (Tris buffer, 50 mM) or pH=6.0 (MES buffer, 50 mM) that contains 1 mM **3**, 1.5 mM GPP, 20 mM $MgCl_2$, 50 µM WelP1 and 5 uM WelU1. The reaction was incubated at 30 °C for 30 mins before quenched by 1 mL EtOAc extraction. The EtOAc crude was dried and dissolved in MeOH for HPLC analysis. For preparative scale reaction for isolating WelU1 and WelU3 enzymatic product, assays were scaled to 10 mL scale proportionally and the crude materials were purified by semi-preparative HPLC.

Table S1. Oligos used for clonings

Oligo #	Oligo Name	Oligo Sequence
1	WelU1-5'	**GATCGGATCC**ATGAAACGAAATTTTATCATTG
2	WelU1-3'	GATCGCGGCCGCTCAAGTTTCAGCTGGTTCG
3	WelU2-5'	**GATCGGATCC**ATGAAGCGAAATTTGATGATTG
4	WelU2-3'	GATCGCGGCCGCTTATGTTCCGGTGGGTTCTG
5	WelU1_5'-trucated	**GATTGGATCC**GCAAGTGCTGTTTCGATTC
6	WelU2_5'-trucated	**GATTGGATCCG**CAGGTGCTACTTTAATTACGATC
7	WelC1-5'	**GATCGGATCC**ATGATTAGAATTATTGCTATTG
8	WelC1-3'	**GATCGCGGCCGC**TTAATTACTACAGATGTTCAAAG
9	WelC3-5'	**GATCGGATCC**ATGACTGCTGAAAACCGACTG
10	WelC3-3'	**GATCGCGGCCGC**TCAAAATGTTGGCAACCAATTTAC

Figure S1. a) In vitro characterization of WelU1/C1/C3 using 1 as a substrate. b) Metal dependency of WelU1 at pH=9. c) Thermostability of WelU1 at pH=9. Chromatographs shown were derived from HPLC analysis, selectively monitored at 280 nm.

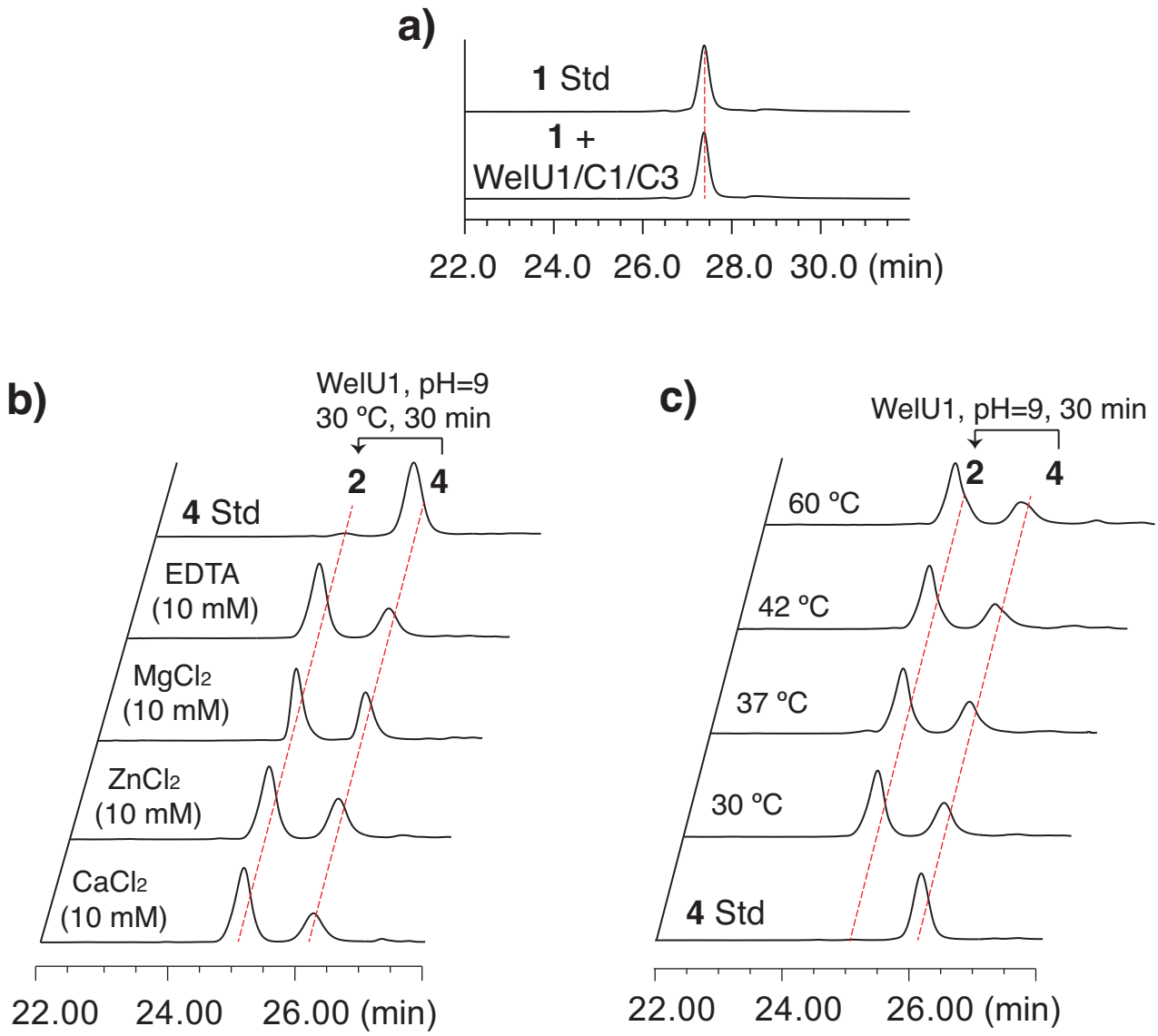


Figure S2. ^{1}H NMR spectral overlay of 12-*epi*-fischerinole U **2** generated enzymatically from **4** with WelU1 with that reported in reference 2.

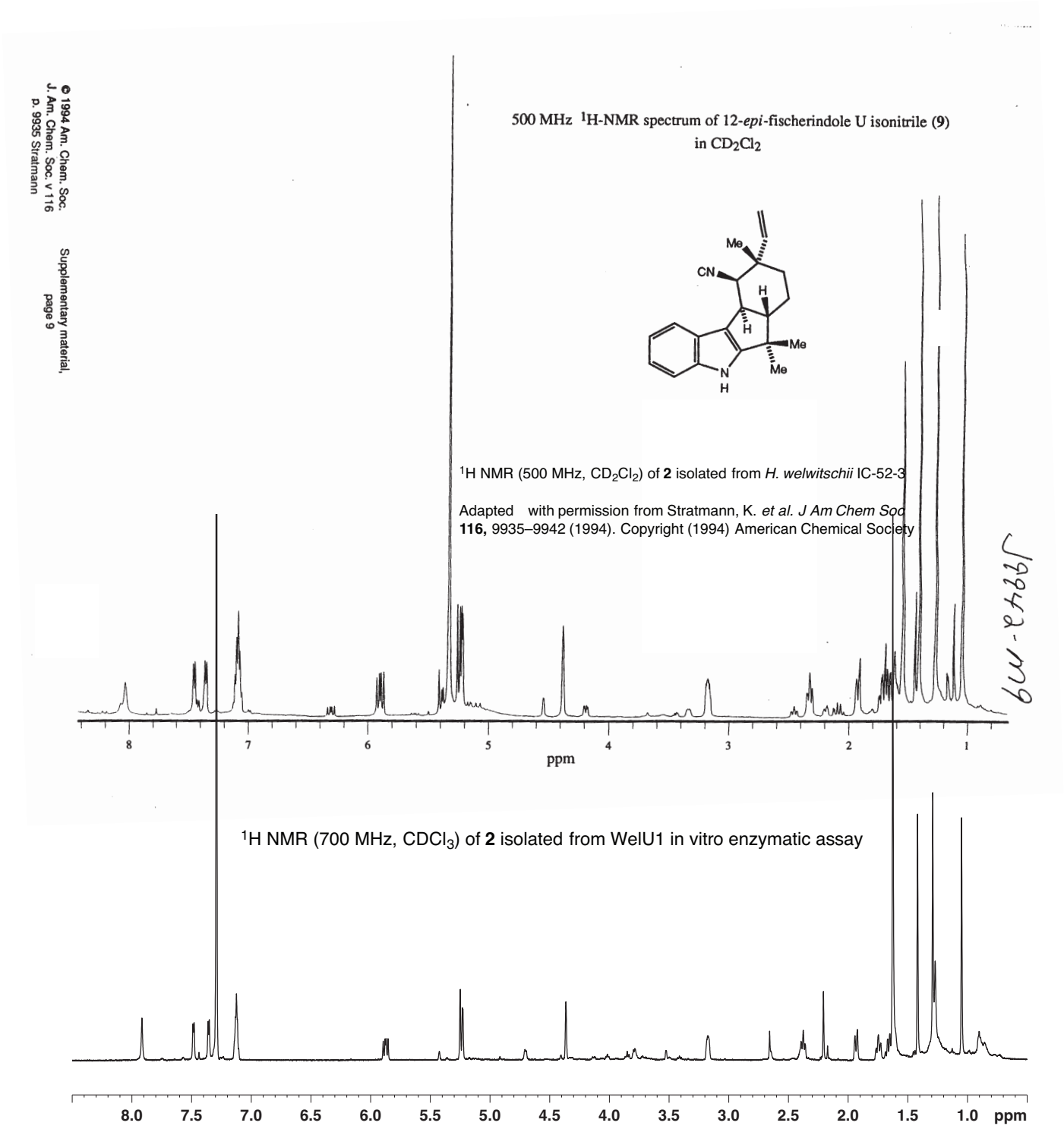


Figure S3. In vitro assay shows compound **5** is not a substrate for WelU1

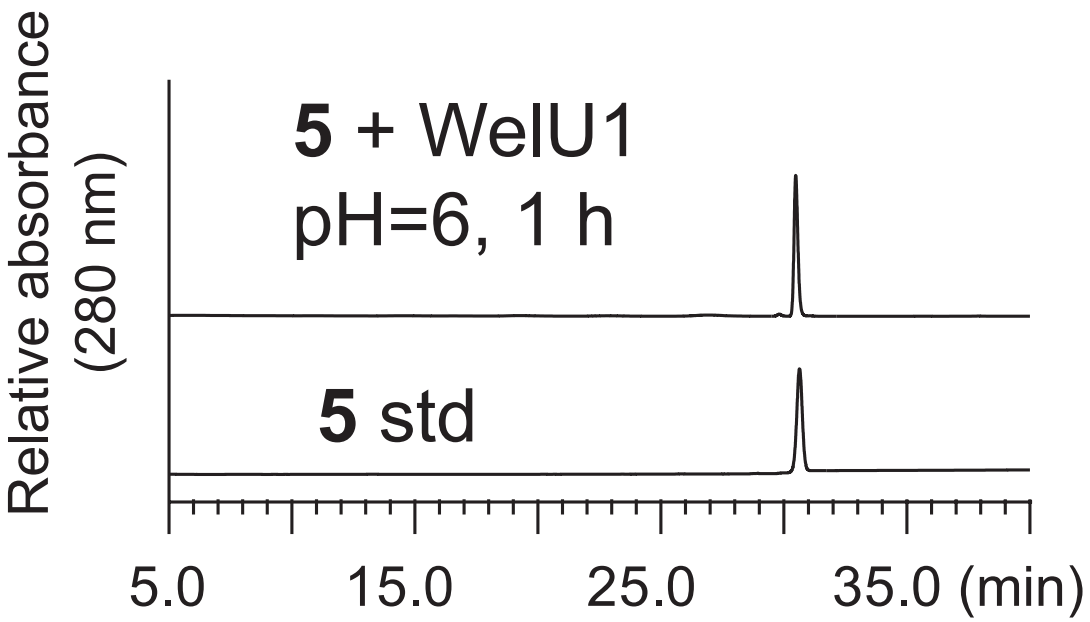
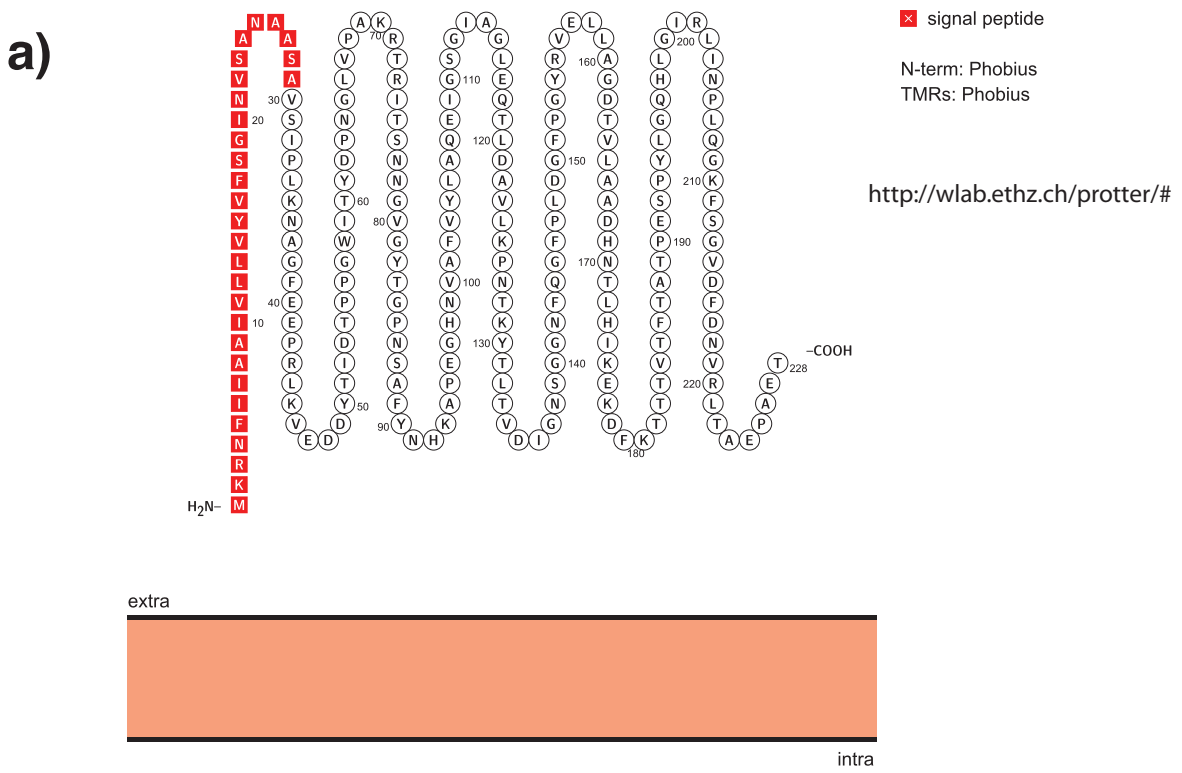
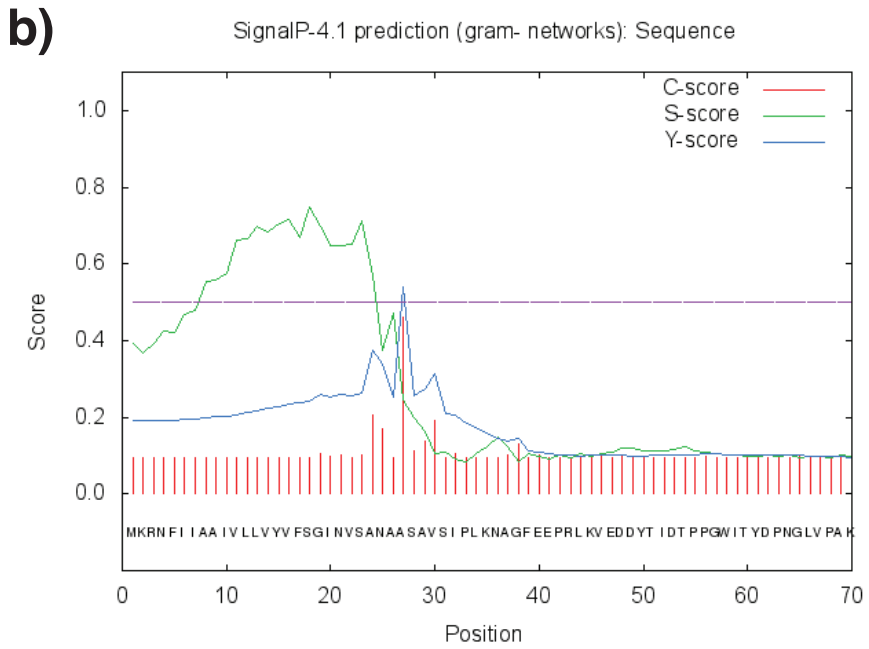


Figure S4. Bioinformatical analysis of WelU1 using (a) Plotter (http://wlab.ethz.ch/protter/start/) and (b) SignalP (http://www.cbs.dtu.dk/services/SignalP/) reveals the potential structural role of N-terminal peptide sequence.

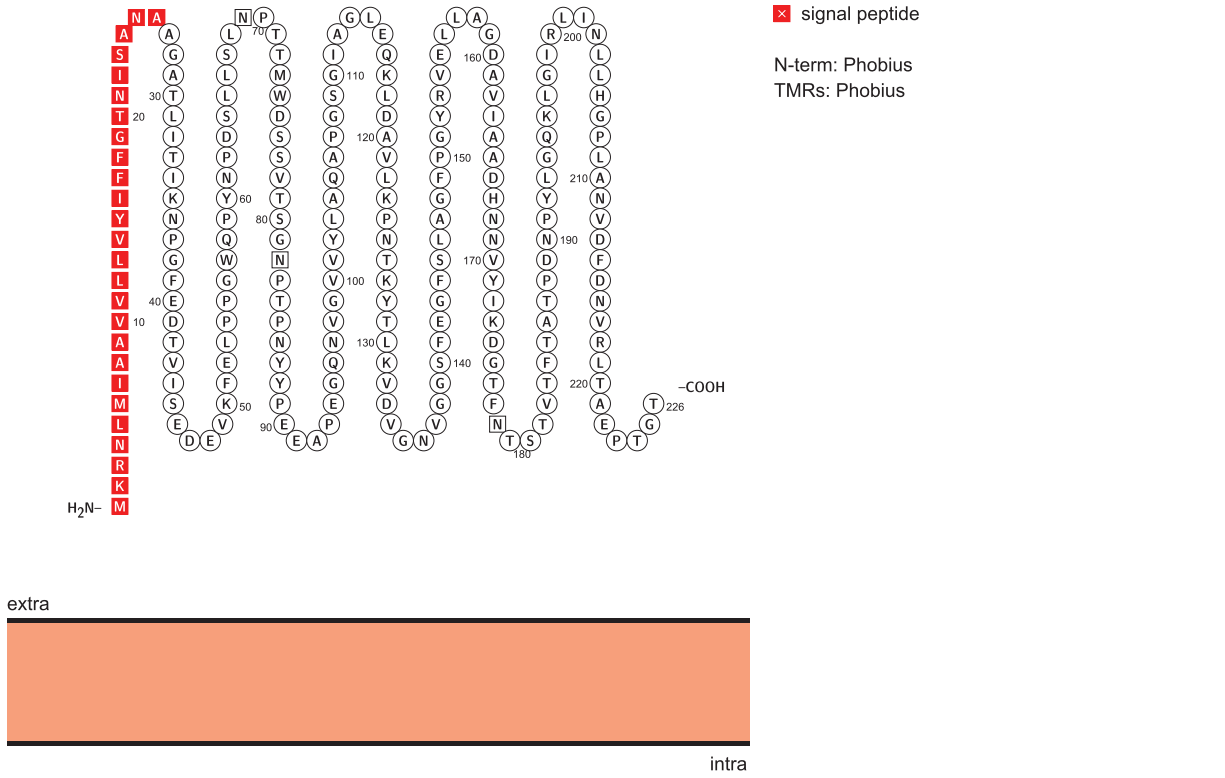




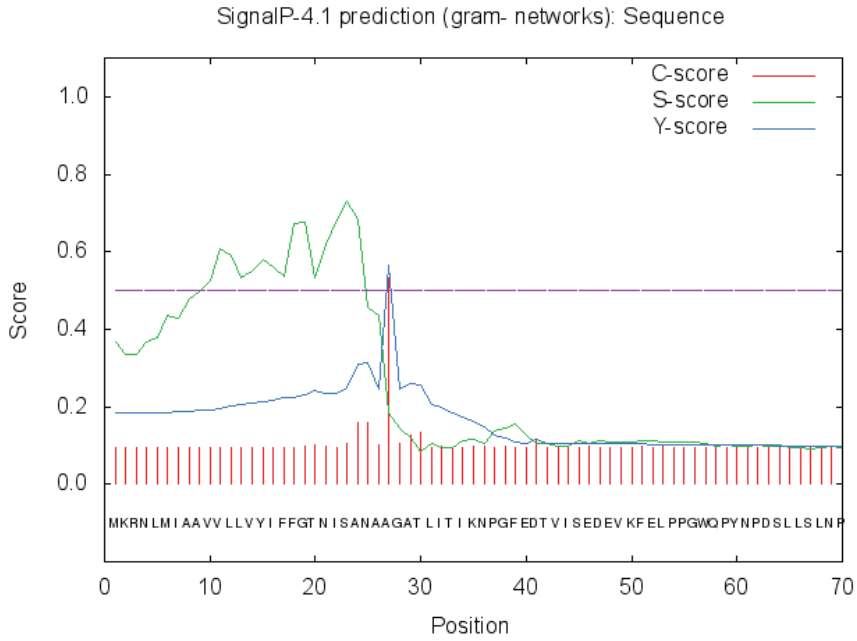

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    27        0.462
  max. Y    27        0.541
  max. S    18        0.749
  mean S     1-26     0.575
       D     1-26     0.554   0.510    YES
Name=Sequence   SP='YES' Cleavage site between pos. 26 and 27: ANA-AS D=0.554 D-cutoff=0.510 Networks=SignalP-TM
```

Figure S5. Bioinformatical analysis of WelU2 using (a) Plotter (http://wlab.ethz.ch/protter/start/) and (b) SignalP (http://www.cbs.dtu.dk/services/SignalP/) reveals the potential structural role of N-terminal peptide sequence.

a)



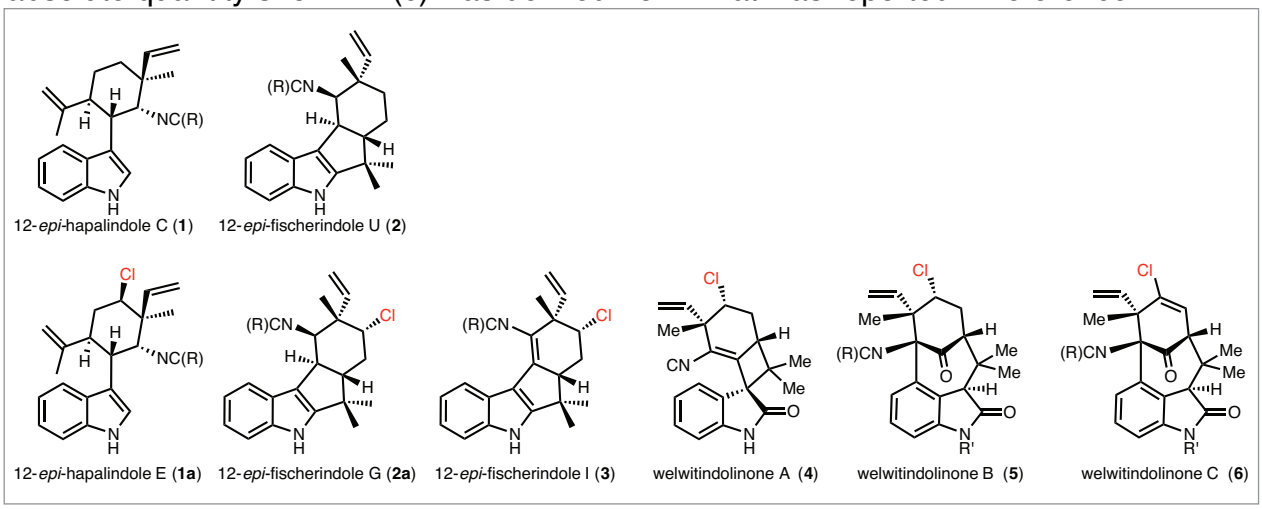
b)



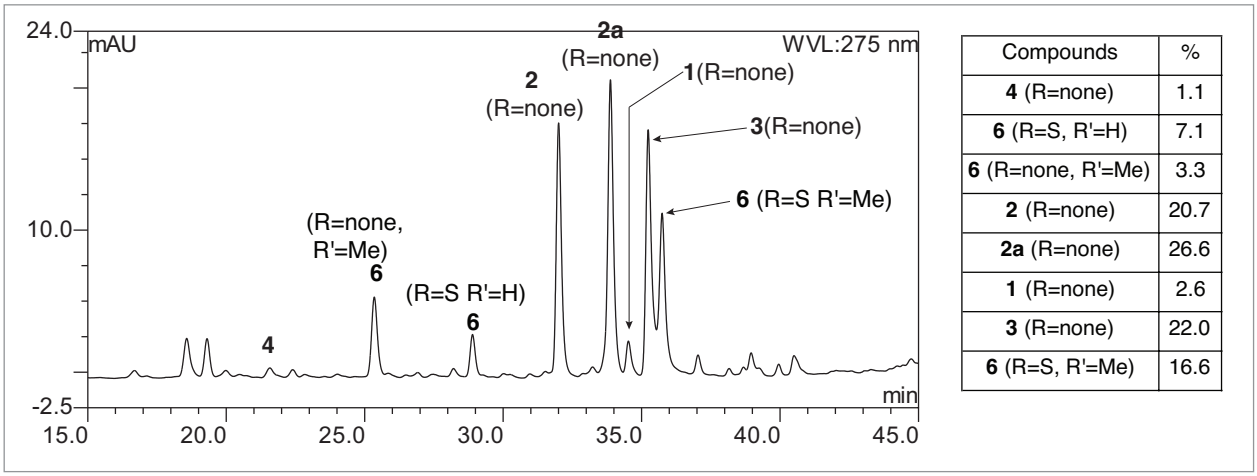

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    27        0.532
  max. Y    27        0.564
  max. S    23        0.731
  mean S     1-26     0.523
       D     1-26     0.549   0.510   YES
Name=Sequence   SP='YES' Cleavage site between pos. 26 and 27: ANA-AG D=0.549 D-cutoff=0.510 Networks=SignalP-TM
```

Figure S6. Summary of relative and absolute quantities of **1** and its biogenetic derivatives **1a** versus **2** and its biogenetic derivatives **2a**, **3**-**6** from (b) *H. welwitschii* UTEX B1830 and (c) IC-52-3. The compounds with number 3-6 are not identical to those described in the main text with the same number. The numbering system (**3**-**6**) here is solely designed for this figure. The relative quantity shown in (b) was derived from the HPLC analysis of *H. welwitschii* UTEX B1830 crude metabolites and the absolute quantity shown in (c) was derived from what was reported in reference 2.

a)



b)



Compounds	%
4 (R=none)	1.1
6 (R=S, R'=H)	7.1
6 (R=none, R'=Me)	3.3
2 (R=none)	20.7
2a (R=none)	26.6
1 (R=none)	2.6
3 (R=none)	22.0
6 (R=S, R'=Me)	16.6

c)

hapalindoles and welwitindolinones isolated from *H. welwitschii* IC-52-3 (47g dry tissue)		
Compound	Name	Quantity isolated
1 (R=none)	12-epi-Hapalindole C isonitrile	11 mg
1 (R=S)	12-epi-Hapalindole C isothiocyanate	1 mg
1a (R=none)	12-epi-Hapalindole E isonitrile	93 mg
1a (R=S)	12-epi-Hapalindole E isothiocyanate	4 mg
2 (R=none)	12-epi-Fischerindole U isonitrile	4 mg
2 (R=S)	12-epi-Fischerindole U isothiocyanate	2 mg
2a (R=none)	12-epi-Fischerindole G isonitrile	4 mg
3 (R=none)	12-epi-Fischerindole I isonitrile	10 mg
4 (R=none)	welwitindolinone A isonitirle	2 mg
5 (R=S, R'=H)	welwitindolinone B isothiocyanate	10 mg
5 (R=S, R'=Me)	N-methylwelwitindolinone B isothiocyanate	5+12 mg
6 (R=S, R'=H)	welwitindolinone C isothiocyanate	14 mg
6 (R=none, R'=Me)	N-methylwelwitindolinone C isonitrile	47mg
6 (R=S, R'=Me)	N-methylwelwitindolinone C isothiocyanate	110mg

Figure S7. ^{1}H NMR spectral overlay of 12-*epi*-hapalindole C **1** generated enzymatically from **4** with WelU3 with that reported in reference 2.

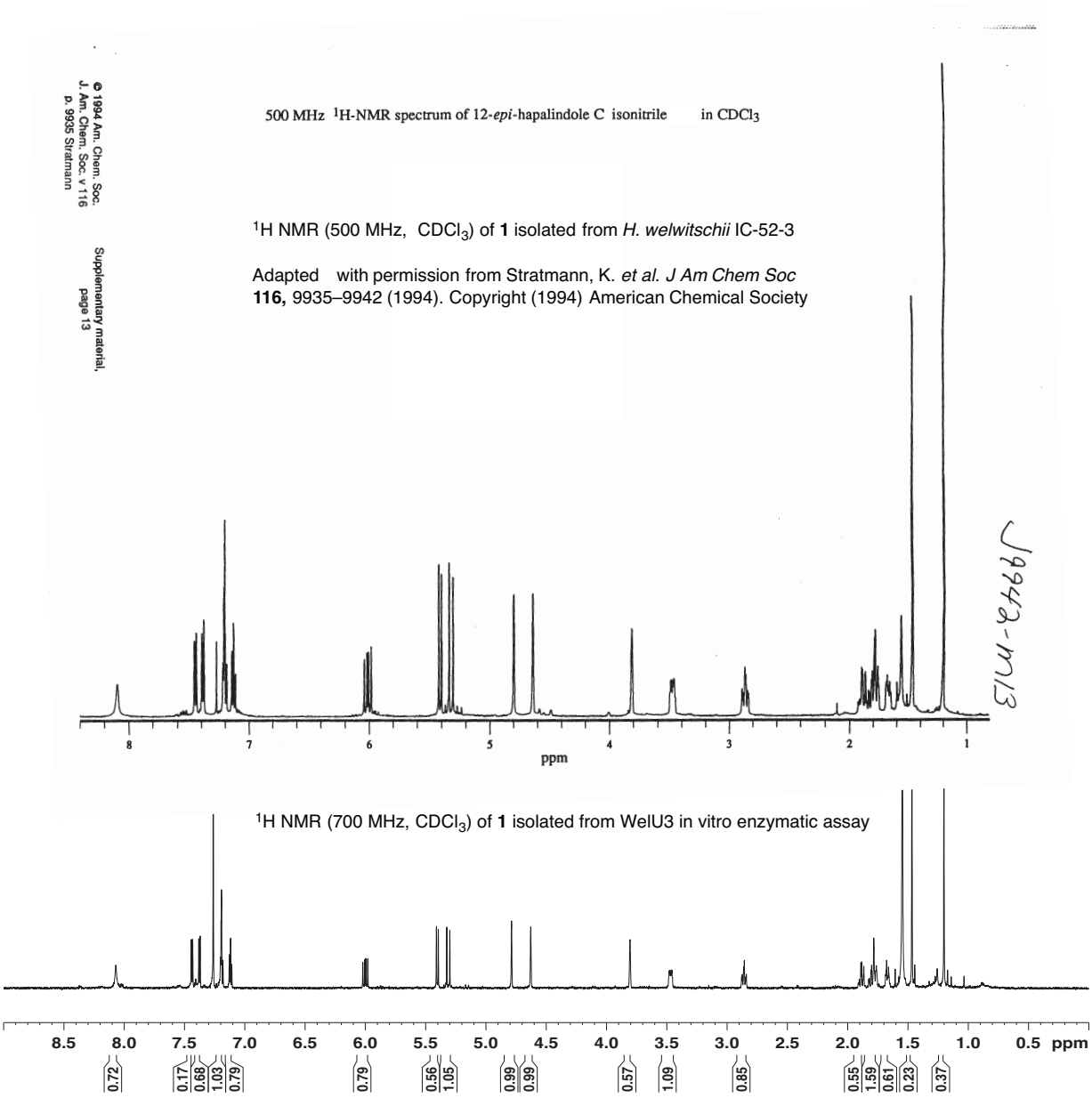


Figure S8. In Vitro characterization of a heterologously expressed and purified AmbU4 in ambiguine biogenesis. Assay was conducted in an identical fashion as described for WelU1 and WelU3 with **4** as the substrate. Product was analyzed by HPLC and monitored at 280 nm. Authenticity of the product was validated by ^{1}H NMR spectral analysis that matched the reported data described in reference 3. .

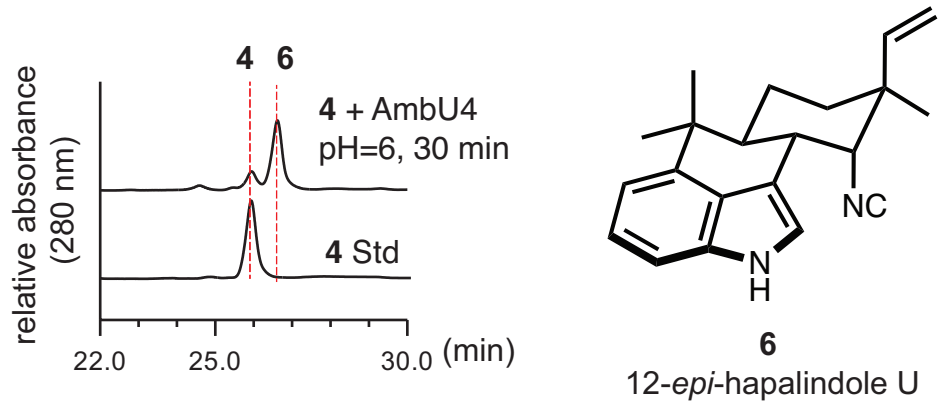


Figure S9. Biomimetic total synthesis of hapalindole-type alkaloids using aza-Prins cyclization as a key step as illustrated in reference 4. Substrate **S1**, when treated with DDQ, gave stereoselective formation of adduct S5 where substituents at C10, C11 and C15 are in an all-equatorial configuration. The reaction is presumed to proceed through the benzylic cation **S2**, which undergoes conformational change to **S3** under thermodynamic control that further leads to imminium ion **S4** for an aza-Prins clyclization.

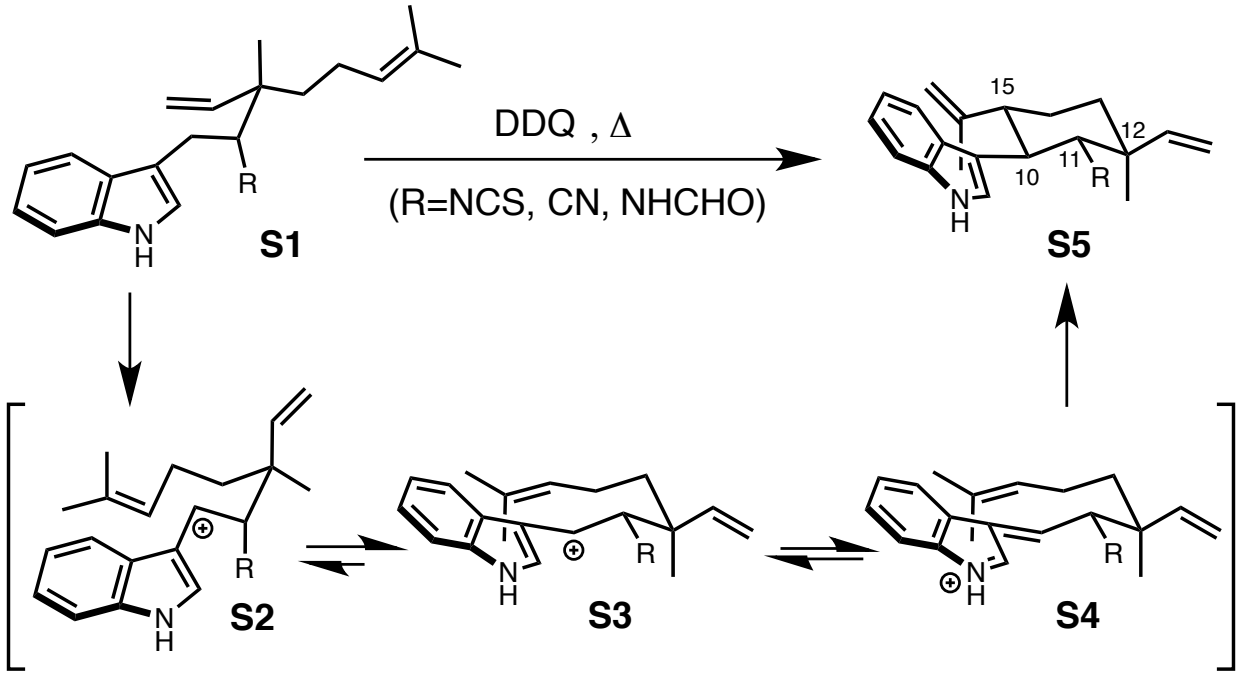
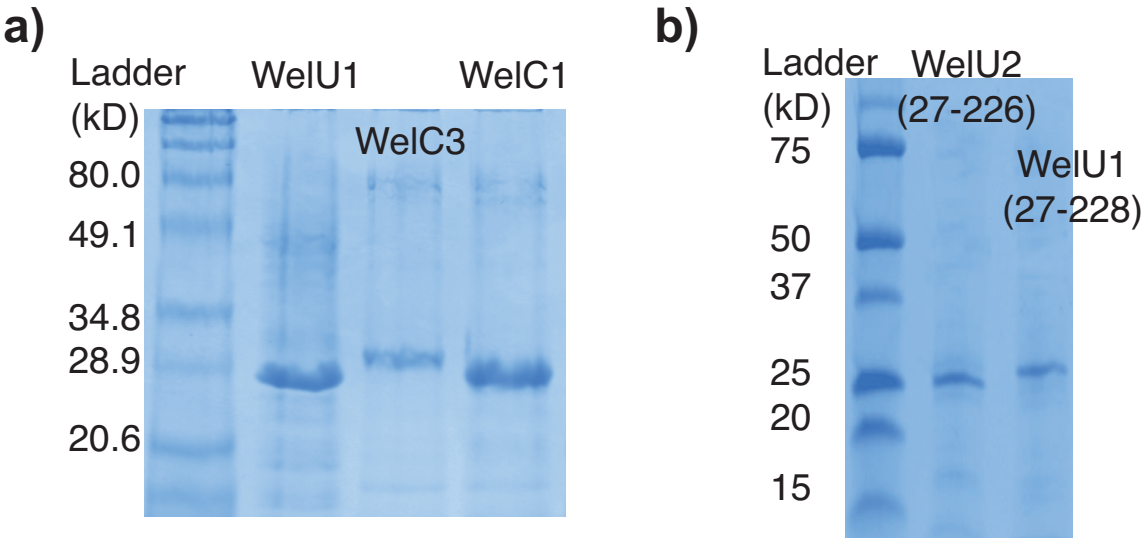


Figure S10. SDS-PAGE of representative proteins heterologously expressed and purified in this work.



Reference Cited:

1. X. Liu, M. L. Hillwig, L. M. Koharudin and A. M. Gronenborn, *Chem. Commun. (Camb.)*, 2016, **52**, 1737-1740.

2. K. Stratmann, R. E. Moore, R. Bonjouklian, J. B. Deeter, G. M. L. Patterson, S. Shaffer, C. D. Smith and T. A. Smitka, *J. Am. Chem. Soc.*, 1994, **116**, 9935-9942.

3. S. Li, A. N. Lowell, F. Yu, A. Raveh, S. A. Newmister, N. Bair, J. M. Schaub, R. M. Williams and D. H. Sherman, *J. Am. Chem. Soc.*, 2015, **137**, 15366-15369.

4. Z. Lu, M. Yang, P. Chen, X. Xiong and A. Li, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 13840-13844.