Supplementary Information (SI) for Green Chemistry. This journal is © The Royal Society of Chemistry 2024

# **Supplementary Information**

Toward safer and more sustainable by design biocatalytic amide-bond coupling

Elisabeth Söderberg <sup>a,b</sup>, Kerstin Felicitas von Borries <sup>c</sup>, Ulf Norinder <sup>d</sup>, Mark Petchey <sup>e</sup>, Ganapathy Ranjani <sup>a,b</sup>, Swapnil Chavan <sup>f</sup>, Hanna Holmquist <sup>g</sup>, Magnus Johansson <sup>h</sup>, Ian Cotgreave <sup>f</sup>, Martin A. Hayes <sup>e</sup>, Peter Fantke <sup>c</sup>, Per-Olof Syrén <sup>a,b</sup>

<sup>a</sup> School of Chemistry, Biotechnology and Health, Science for Life Laboratory, KTH Royal Institute of Technology, Tomtebodavägen 23 171 65, Solna, Sweden

<sup>b</sup> School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Fibre and Polymer Technology, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

<sup>c</sup> Quantitative Sustainability Assessment, Department of Environmental and Resource Engineering, Technical University of Denmark, Bygningstorvet 115, 2800 Kgs. Lyngby, Denmark

<sup>d</sup> Department of Computer and Systems Sciences, Stockholm University, 164 25 Kista, Sweden

<sup>e.</sup> Discovery Sciences, BioPharmaR&D, AstraZeneca, 431 83 Mölndal, Sweden

<sup>f</sup> The Research Institute of Sweden RISE, Chemical and pharmaceutical toxicology, 151 36 Södertälje, Sweden

<sup>g.</sup> IVL Swedish Environmental Research Institute, Aschebergsgatan 44, 411 33 Göteborg, Sweden

<sup>h.</sup>Early CVRM Medicinal Chemistry, BioPharmaR&D, AstraZeneca, 431 83 Mölndal, Sweden

# **Table of Contents**

Supplementary Notes	4
Literature study ATP-dependent amide bond synthetases	4
USEtox average CF:s	4
Table S1. Average CF values for safechems and their corresponding amides	4
Table S2. Percentile of average CF values for the subset of safechems used in the experimental coupling compare the percentile of the average CFs for the 16 acids and 17 amines with lowest CF values.	d to 4
Methods	5
Protein expression and purification	5
Thermostability	5
UPLC-MS	5
Synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide	5
Synthesis of 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide and N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide	5
Table S3. Sequence identity (%) between McbA and its ancestors A1-A4	7
Table S4. Melting temperature ( $T_m$ ) of McbA and ancestors A1-A4 measured by nano differential scanning fluorimetry.	7
Table S5. Protein expression levels of McbA and ancestors.	7
Figure S1. Crude protein content of McbA and ancestor A1-A4.	7
Figure S2. Thermal unfolding by McbA and ancestors A1-A4	8
Figure S3. Phylogenetic tree of McbA used to compute ancestral sequences	8
Figure S4. Alignment used for the phylogenetic tree construction.	12
Figure S5. Alignment of McbA and ancestors A1-A4	12
Figure S6. <sup>1</sup> H-NMR of synthesised 3-Hydroxy-N-(3-phenylpropyl)benzamide (standard)	13
Figure S7. <sup>13</sup> C-NMR of synthesised 3-Hydroxy-N-(3-phenylpropyl)benzamide (standard).	13
Figure S8. <sup>1</sup> H-NMR of the enzymatic synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide (crude NMR)	14
Figure S9. <sup>13</sup> C-NMR of the enzymatic synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide (crude NMR)	14
Figure S10 <sup>13</sup> C-NMR (left) and <sup>1</sup> H-NMR (right) of 3-hydroxybenzoic acid, 3-phenyl-1-propylamine, and the enzymatic syr of 3-Hydroxy-N-(3-phenylpropyl)benzamide (crude NMR)	nthesis 15
Figure S11. <sup>1</sup> H-NMR of synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide (standard)	16
Figure S12. <sup>13</sup> C-NMR of synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide (standard)	16
Figure S13. <sup>1</sup> H-NMR of biocatalytic synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide	17
Figure S14. <sup>13</sup> C-NMR of biocatalytic synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide	17
Figure S15. DEPT spectra of biocatalytic synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide	18
Figure S16. <sup>1</sup> H-NMR of synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide (standard).	19
Figure S17. <sup>13</sup> C-NMR of synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide (standard)	19
Figure S18 <sup>1</sup> H-NMR of biocatalytic synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide	20
Figure S19 <sup>13</sup> C-NMR of biocatalytic synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide	20
Figure S20. DEPT spectra of biocatalytic synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide	21
Figure S21. 305 nm UV absorption calibration curve of different 3-hydroxybenzoic acid concentrations.	22

Figure S22. HPLC-MS chromatogram of a sample taken from the upscaled biocatalytic synthesis of 3-Hydroxy-N-(3- phenylpropyl)benzamide	22
Figure S23. HPLC-MS chromatogram of 3-Hydroxy-N-(3-phenylpropyl)benzamide standard	23
Figure S24. 305 nm UV absorption calibration curve of different 3-acetylbenzoic acid concentrations	23
	23
Figure S25. HPLC-MS chromatogram of a sample taken from the upscaled biocatalytic synthesis of 3-acetyl-N-(4- (hydroxymethyl)phenethyl)benzamide	24
Figure S26. HPLC-MS chromatogram of 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide standard	24
Figure S27. 305 nm UV absorption calibration curve of different benzoic acid concentrations	25
Figure S28. HPLC-MS chromatogram of a taken from the upscaled biocatalytic synthesis of N-(2-(piperidin-1- ylsulfonyl)benzyl)benzamide	25
Figure S29. HPLC-MS chromatogram of N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide standard	26
Table S6. Derivation of USEtox input parameters' mean and standard deviation from in-silico model prediction and uncertainty.	27
Table S7. Overview of uncertainty factors applied onto the model-specific uncertainty	28
Figure S30.USEtox characterization factors vs <i>in-silico</i> screening scores of safechems and subset of filtered out aroma and amines.	tic acids 28
Figure S31. Predicted toxicity and environmental score of detected amides from the experimental coupling considerir subset of possible safechems in comparison to all possible amides from the whole safechem panel.	ıg only a 28
Figure S32. Toxic effects coupled to the molecular weight of amines and acids	29
Figure S33. Scatter plot of the USEtox CFs of amides with <i>in silico</i> hazard score equal to or lower than ten (figure 3) co to the sum of the CFs of their constituent acid and amine moieties.	mpared 29
Table S8. Conversion by McbA (in %)	30
Table S9. Conversion by A1 (in %).	31
Table S10. Conversion by A2 (in %)	32
Table S11. List of safechem amines from <i>in silico</i> screening	33
Table S12. List of safechem acids from in silico screening	35
Table S13. The exact masses of detected and potential amides in the experimental coupling in figure 4.	37
Notes and references	

## **Supplementary Notes**

#### Literature study ATP-dependent amide bond synthetases

Many of the characterised ABS enzymes have been identified in genome mining projects of gene clusters responsible for the biosynthesis of various medically relevant secondary metabolites. NovL, (1) CloL, (2) SimL, (3) and CouL (4,5) are involved in the biosynthesis of aminocoumarin antibiotics novobiocin, clorobiocin, simocyclinone D8, and coumermycin A1 (figure 2A, panel C). Another group of ABS-reminiscent enzymes that share sequence motifs and similar reaction mechanisms to ABS-enzymes, are Ann1, (6) AsuD1 (7) and ORF33 (figure 2A, panels B, D, E) (8,9). These enzymes are involved in the synthesis of 2-amino-3-hydroxycyclopent-2-enone (C5N) and tailoring of polyketides. Recently, a group of CfaL enzymes responsible for the ligation of coronafacic acid (CFA) and coronamic acid (CMA) in the bacterial plant hormone-mimic phytotoxin coronatine (COR) was characterised and screened against different acyl and amine substrates (figure 2A, panel F). The enzymes accepted a broad range of both acyl and amine substrates; aliphatic, aryl, and heteroaryl carboxylic acids, and non-proteogenic amino acids. (10) The enzymes also managed to utilize carboxylic acids carrying other reactive functionalities, groups that would have needed to be protected if the substrate were to be used in chemical synthesis; again, emphasizing the potential of biocatalysts in amide bond synthesis.

#### **USEtox average CF:s**

The average values of the CFs of the subset of arbitrarily selected safechems that were used in the experimental coupling did not differ much compared to the average values of all safechems (table S1). The average human CFs of the acids used were slightly lower than those of all safechem acids, while the opposite was true for the amines. For ecotoxicity, amines and acids in the subset were predicted slightly more toxic compared to all safechems. As for the amides made by the subset safechems, their average CFs did not differ much from the average of all amides made by safechems passing the established threshold (figure 3, table S1). The percentiles of the average CF values of the subset were between 44.3 to 55.4 %. The average CF percentile value of an equally large subset of safechem acids and amines could go as low as 19.9 and 4.7%, respectively (table S2). If we had implemented USEtox after the *in silico* filtering and before the experimental coupling, we could have reduced the human and ecotoxicity of the panel used in the experimental coupling even further.

Table S1. Average CF values for safechems and their corresponding amides. For reference, the average CFs are presented separately for all safechems, their responding amides and for the subset that was used in the experimental coupling, respectively. Amides (row 5) are the amides passing the established threshold in the *in silico* filtering (figure 3).

Table S2. Percentile of average CF values for the subset of safechems used in the experimental coupling compared to the percentile of the average CFs for the 16 acids and 17 amines with lowest CF values. Data is shown for each of the four categories.

	Freshw	ater ecotoxicity	Human toxicity		
		(CF's)	(CF's)		
Туре	Air	Freshwater	Air	Freshwater	
Safechem acids	-1.34	0.34	-4.07	-4.48	
Safechem acids					
used in the	1 10	0.42	4.15	4.60	
experimental	mental -1.18 0.43	0.43	-4.15	-4.60	
coupling					
Safechem amines	-0.72	1.16	-3.97	-4.17	
Safechem amines					
used in the		1.15	204	4.12	
experimental	-0.67	1.15	-3.94	-4.13	
coupling					
All amides	-0.11	1.50	-3.02	-3.50	
Amides from					
safechems used					
in the	0.01	1.64	-3.05	-3.53	
experimental					
coupling					

	Percentile of average CF							
	Freshwater	ecotoxicity	Human toxicity					
Туре	Air	Freshwater	Air	Freshwater				
Subset of								
16 acids	50.9	46.0	52.2					
used in the	50.8	46.9	52.2	41.4				
coupling								
16 acids								
with the	10.0	23.9 26.4	22.0 25.4	22.0 26.4	22.0	22.0	25.4	20.8
lowest CF	19.9		23.9 20.4	30.8				
value								
Subset of								
17 amines	47.4	44.2	54.4	AE A				
used in the	47.4	44.5	54.4	40.4				
coupling								
17 amines								
with the	5.2	4.7 6.3	4.7 6.3	4.7 6.3	6.3	47 62	6.1	
lowest CF	5.3					0.1		
values								

# Methods

#### Protein expression and purification

The four consensus sequences were codon optimised for expression in Escherichia Coli by GeneArt (Thermo Fisher Scientific). Kanamycin resistance gene and N-terminal His-tag were included in all constructs, and each construct was delivered in pET-28a(+) plasmids. BL21(DE3) competent cells were transformed with approximately 2 ng plasmid, and selection was made by growing the cells on 30 µg/mL kanamycin lysogeny broth (LB) agar plates. For protein expression, overnight cultures grown in LB supplemented with 30 µg/mL kanamycin were added to LB with 30 µg/mL kanamycin to a start OD600 of 0.1. When the OD600 reached 0.6, 1 mM IPTG was added. McbA, A3, and A4 were set to grow overnight at 16 °C and 180 rpm, while A2 and A3 at 30 °C. The cells were harvested by centrifugation at 5,000 g for 20 min, and pellets were resuspended in 50 mM phosphate buffer (pH 7.5) with 10 % glycerol (w/v), 300 mM NaCl, and 40 mM imidazole. The cell suspensions were kept on ice while being sonicated at 30s x 3 bursts with 1.5 minute breaks using Branson Sonifier 450 (35% duty cycle and output control 5). Afterwards, the sonicated cell suspensions were centrifuged for 30 min at 4 °C and 14,000 g. The supernatants were injected into the ÄKTA explorer protein purification system (Cytiva) at 4 ml/min, equipped with a 5 mL HP-HisTrap column (Cytiva) equilibrated with ten column volumes of binding buffer. After washing the column with 20 column volumes of binding buffer, the proteins were eluted by a gradient going from 40 mM to 300 mM imidazole over 20 column volumes. Elution fractions were analysed by Invitrogen NuPAGE 4-12% Bis-Tris gel (Thermo Fisher Scientific), and desalted with PD10 columns (Sigma-Aldrich) equilibrated with 50 mM phosphate buffer (pH 7.5) and 300 mM NaCl. Concentration of protein samples was made with MacroSep 10 kDa cut-off centrifugation filters (Pall Laboratory). Protein concentration was measured with NanoDrop. 80% glycerol solution in pH 7.5 phosphate buffer with 300 mM NaCl was added to the protein samples to a final concentration of 10%. The proteins were aliquoted to 1.5 mL microcentrifuge tubes, snap-frozen in liquid nitrogen, and stored at -80 °C.

#### Thermostability

The thermostability of the proteins was assessed by nano differential scanning fluorimetry using Prometheus NT.48 nanoDSF (NanoTemper Technologies). Samples of 2 mg/mL protein solution were prepared in glass capillaries. The fluorescence of the proteins at 330 and 350 nm was measured at a temperature gradient going from 20 to 95 °C at 1 °C/min. The derivate of the ratio between the two measured wavelengths over the temperature interval was used to determine the melting temperature of the proteins.

#### UPLC-MS

The reactions were quenched with 30  $\mu$ L acetonitrile supplemented with 1 mg/mL 1,2,3-tribromobenzene, and the plates were sealed by Velocity11's PlateLoc. The plates were shaken in a short burst followed by centrifugation at 4000 g for 10 min. Waters Acquity UPLC equipped with a photodiode array detector, and a 3100 mass spectrometer was used for analysis. 3  $\mu$ L samples were injected into an Acquity UPLC HSS C18 column (1.8  $\mu$ m, 2.1 mm X 50 mm). The mobile phases were prepared by adding 0.5 mL of acidic solution (pH 3 solution consisting of 126.3 g water, 151.8 g formic acid, and 21.8 g ammonium hydroxide solution 25-30%) to 1 L of water (mobile phase A), and 1 mL of the acidic solution was added to 1 L 95% acetonitrile (mobile phase B). The mobile phase went from 10% to 99% B in 2 minutes. The conversions of the reactions were measured as the amide peak area percentage of the total amide and acid peak area using MassLynx (Waters).

#### Synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide

3-hydroxybenzoic acid (0.511 g, 3.70 mmol) was dissolved in DCM (10 mL). To this, 1H-benzo[d][1,2,3]triazol-1-ol hydrate (0.623 g, 4.07 mmol) and 3-phenylpropan-1-amine (0.526 ml, 3.70 mmol) were added. Following this, diisopropylmethanediimine (0.573 ml, 3.70 mmol) was added to the mixture and the reaction was mixed for 4 h. Upon completion of the reaction, the diisopropylurea was filtered away and the DCM layer was washed with saturated bicarbonate solution (1×15 mL) and 2 M HCl (1×15 mL). The organic layer was dried over anhydrous sodium sulfate and the solvent was then removed under reduced pressure. The crude product was purified by flash column chromatography (Ethyl acetate: Heptane gradient 5 – 50% over 12 CV) to yield the desired product as a white solid in 40% yield. <sup>1</sup>H NMR (500 MHz, DMSO)  $\delta$  9.60 (s, 1H), 8.37 (t, *J* = 5.6 Hz, 1H), 7.34 – 7.14 (m, 8H), 6.92 – 6.86 (m, 1H), 3.25 (td, *J* = 7.1, 5.6 Hz, 2H), 2.65 – 2.58 (m, 2H), 1.81 (qd, *J* = 7.7, 6.4 Hz, 2H). <sup>13</sup>C NMR (126 MHz, DMSO)  $\delta$  166.3, 157.3, 141.8, 136.2, 129.2, 128.3, 128.3, 125.7, 117.9, 117.6, 114.2, 38.8, 32.6, 30.9.

#### Synthesis of 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide and N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide

In a round bottom flask the acid (1.0 equiv.), EDC (1.5 equiv.), HOBt (1.0 equiv.), the amine (1.2 equiv.), and DCM: DMF (10:1) were added. The reaction mixture was stirred at room temperature overnight, and diluted with DCM. The organic phase was washed with saturated NaHCO<sub>3</sub>, and brine, dried over MgSO<sub>4</sub>, filtered, and concentrated in vacuo. The crude was purified by preparative

TLC (2%-10% MeOH-CHCl<sub>3</sub>) afford corresponding to the amide compound. **3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide**:  $R_f = 0.44$  (MeOH/CHCl<sub>3</sub> (10:90)), Colourless solid, Yield (16 mg, 44%) <sup>1</sup>H NMR, 400 MHz, DMSO-*d*<sub>6</sub>: 8.80 (t, J=5.6 Hz, 1H), 8.4 (s, 1H), 8.08 (t, J=8.0 Hz, 2H), 7.62 (t, J=7.8 Hz, 1H), 7.22 (dd, J=16, 8.0 Hz, 2H), 7.62 (t, J=7.8 Hz, 1H), 7.22 (dd, J=16, 8.0 Hz, 2H), 7.62 (t, J=7.8 Hz, 1H), 7.22 (dd, J=16, 8.0 Hz, 2H), 7.62 (t, J=7.8 Hz, 1H), 7.22 (t, J=7.8 Hz, 1H), 7.24 (t, J=7.8 Hz, 4H), 5.12 (t, J=5.7 Hz, 1H), 4.45 (d, J=5.7 Hz, 2H), 3.49 (dd, J=14.0, 6.6 Hz, 2H), 2.84 (t, J=7.4 Hz, 2H), 2.63 (s, 3H). <sup>13</sup>C NMR, 400 MHz, DMSO-*d*<sub>6</sub>: 197.7, 165.4, 140.4, 137.8, 136.8, 135.0, 131.8, 130.8, 128.9, 128.4, 126.8, 126.6, 62.8, 41.1, 34.8, 26.9. N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide: R<sub>f</sub>= 0.5 (MeOH/CHCl<sub>3</sub> (2:98)), Colourless solid, Yield (5.0 mg, 34%) <sup>1</sup>H NMR, 400 MHz, DMSO-*d*<sub>6</sub>: 9.07 (t, *J*=5.5Hz, 1H), 7.95 (d, *J*=7.2Hz, 2H), 7.84 (d, *J*= 7.7 Hz, 1H), 7.65 (t, *J*=7.5 Hz, 1H), 7.58 (t, J=7.1Hz, 1H), 7.53-7.47 (m, 4H), 4.85 (d, J=5.9 Hz, 2H), 3.10 (t, J=5.2 Hz, 4H), 1.57 (br, 4H), 1.47 (d, J=4.3Hz, 2H) <sup>13</sup>C NMR, 400 MHz, DMSO-*d*<sub>6</sub>: 166.6, 138.5, 134.5, 134.0, 133.2, 131.5, 129.8, 128.4, 127.7, 127.3, 127.1, 45.7, 39.6, 24.9, 23.1

Table S3. Sequence identit	v (%	) between McbA	and its	ancestors A1-A4.

	McbA	A1	A2	A3	A4
McbA	100.0	70.0	62.4	50.4	40.0
A1	70.0	100.0	85.2	65.0	48.8
A2	62.4	85.2	100.0	75.4	54.8
A3	50.4	65.0	75.4	100.0	66.8
A4	40.0	48.8	54.8	66.8	100.0

Table S4. Melting temperature  $(T_m)$  of McbA and ancestors A1-A4 measured by nano differential scanning fluorimetry.

Enzyme	T <sub>m</sub> 1 (°C)	T <sub>m</sub> 2 (°C)
McbA	48.1	
A1	73.1	
A2	66.0	79.3
A3	43.8	73.3
A4	67.9	

Table S5. Protein expression levels of McbA and ancestors

Enzyme	Approximate mg protein per g pellet
McbA	10
A1	7
A2	8
A3	14
A4	8



Figure S1. Crude protein content of McbA and ancestor A1-A4. Samples were normalized by OD600, two dilutions of each sample were loaded onto the gel. The molecular weight marker is SeeBlue Plus2.

Melting temperature McbA and ancestors



Figure S2. Thermal unfolding by McbA and ancestors A1-A4, measured by change in fluorescence ratio at wavelengths 350 and 330 nm by Prometheus NT.48 nanoDSF (NanoTemper Technologies).



Figure S3. Phylogenetic tree of McbA used to compute ancestral sequences. The numbers shown on the nodes are the bootstrap values (1000 bootstrap).

		1	10	20
AGL76720.1 Marinactinospora thermotolerans		MGYA	RRVMDGIGEVAV	
WP 051713392.1 Actinoalloteichus cyanogriseus		MDYV	RRVLAGIAPODDVL	VSEN
WP_073481158.1_Streptoalloteichus_hindustanus		MGYL	HRVVEGLKANAGGEAL	VSAD
WP_210842583.1_Nocardiopsis_spB62			MESAGDRNVI	<b>V</b> NGA
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	MRP	MGY <mark>I</mark>	SQIVHGVESAGDRDVI	<b>V</b> DGP
WP_178378553.1_Nocardiopsis_flavescens		MGY <mark>V</mark>	SRIVRGLESIGTKEAV	VLDG
WP_179079716.1_Streptomyces_rectiverticillatus	MSVA	ALPFRSF <mark>V</mark>	DDILDALGADPAREAL	VHQG
WP_208608527.1_Amycolatopsis_pretoriensis		<mark>M</mark>	DDILDSLEAAPEREAL	<b>∨</b> HYG
WP_150246336.1_Streptomyces_albofaciens	MPDT	AHDFRTY <mark>V</mark>	EENLDALGADPERVAL	VHQG
WP_053802994.1_Streptomyces_rimosus	MPDT	AHDFRTY <mark>V</mark>	EENLDALGTDPGRVAL	VHQG
WP_106677009.1_Streptosporangium_nondiastaticum	MTAT	AQQFRTY <mark>V</mark>	ERNLDILGADPTRTAV	<b>∨</b> HHD
WP_184939538.1_Kitasatospora_kifunensis	MPGD	TQQFRSY <mark>V</mark>	EENLDALAQWPQREAV	<b>V</b> YGD
WP_150485975.1_Streptomyces_nitrosporeus	MPTPV	TPAFRNY <mark>V</mark>	EDILDGLRNTPGREIL	VA.GERGE
WP_204453063.1_Actinokineospora_baliensis	MRSRPWEGAAMSAA	APEFRNY <mark>V</mark>	EDILDGLSGQEVV	<b>V</b> SGD
WP_121391275.1_Actinokineospora_cianjurensis	MSTT	APGFRNY <mark>V</mark>	EDILDGLSGPEVV	<b>V</b> SGD
WP_075137134.1_Actinophytocola_xinjiangensis		MPVFHNY <mark>V</mark>	EDVLDGLSAEPGREVL	<b>V</b> SGE
WP_025359162.1_Kutzneria_albida	MTSSTAG	ATGSANY <mark>V</mark>	ETILRGLSRDSDHIAF	HLE.DG
WP_083474567.1_Pseudonocardia_spAL041005-10	MSADLDSATAPRNG.AATQS	GAALHSDV	EDLLDALAADRGREAL	VHRD
WP_208819870.1_Pseudonocardia_alni	MSVDPDSATAPRNG.AATQS	GAALHSDV	EELLDALAADRGREAL	VHRD
WP_083658735.1_Pseudonocardia_autotrophica	MSADLDSATAPRHG.AATQS	GAALHSDV	EELLDALAADRGREAV	VHRD
WP_068/981/5.1_Pseudonocardia_spHH130630-0/	M	IGIGARIDV	EELLAALAADPGREAL	VFGE
WP_043911969.1_Kitasatospora_griseola	MP	DAEFVSGP	AMVLRSLAADPDRQAV	VDA.AG
WP_11/489816.1_Kitasatospora_xanthocidica	MP	DAEFVSGP	ALILQILAADPDRQAV	VIA.DG
MD 068523009 1 Mauhamumalla humasinasaluana	MS	MDDT UDE	DIACDILLCCLAAVDDCVAV	VDA.DG
WP_0000220000.1_ISukamurella_cyrosinosorvens		MDDIURE	PLACDILLCSLAAYPDPVAV	HVIGADGD
WP_120190109.1_ISukamurella_paurometabora		MDDIHRE	PLACDILLCSLCAVSEPTAV	HVTGPOGD
WP_006543910 1 Frankia en FIN1f	MSPG	TREADDIVEC	AFI PDI I VCAL PPDDAKDAT	NVIGEQGD
WP 161991522 1 Natronorubrum aibiense	MSVOT	T.P	SILL TTTISPRER TAL	RD ET
WP 007190185.1 Haloarcula californiae	MLOOS	PPTIKOLY	ESVLERYESAP AV	TELLGD
oo.iooioo.iouiouiouiu_ouiiioiniue			·····	

AGL76720.1_Marinactinospora_thermotolerans
WP_051713392.1_Actinoalloteichus_cyanogriseus
WP_073481158.1_Streptoalloteichus_hindustanus
WP_210842583.1_Nocardiopsis_spB62
PWV44261.1_Nocardiopsis_spL17-MgMaSL7
WP_178378553.1_Nocardiopsis_flavescens
WP 179079716.1 Streptomyces rectiverticillatus
WP 208608527.1 Amycolatopsis pretoriensis
WP 150246336.1 Streptomyces albofaciens
WP 053802994 1 Streptomyces rimosus
WP 106677009.1 Streptosporangium nondiastaticum
WP 184939538 1 Kitasatospora kifunensis
WP 150485975 1 Streptomyces pitrosporeus
WP 204453063 1 Actinokineospora baliensis
MP_204455005.1_ACCINCKINEOSPOIA_DAITENSIS
WP_121391275.1_Actinokineospora_cianjurensis
WP_0/513/134.1_Actinophytocola_xinjiangensis
WP_025359162.1_Kutzneria_albida
WP_083474567.1_Pseudonocardia_spAL041005-10
WP_208819870.1_Pseudonocardia_alni
WP_083658735.1_Pseudonocardia_autotrophica
WP_068798175.1_Pseudonocardia_spHH130630-07
WP_043911969.1_Kitasatospora_griseola
WP_117489816.1_Kitasatospora_xanthocidica
OKJ05654.1_Kitasatospora_spCB01950
WP 068523008.1 Tsukamurella tyrosinosolvens
WP 126196109.1 Tsukamurella paurometabola
WP 068744906 1 Tsukamurella pseudospumae
WP 006543910 1 Frankia sp. EUN1f
WP 161991522 1 Natronorubrum aibiense
MP_101991922.1_MattonordDrum_atDiense
wF_007190185.1_Haloarcula_callfolliae

30	40	50	eo	70
GSVTGARLRHOVRLLAHA	LTEAGIPPGR	GVACLHANTW	RAIALRLAVC	AIGCHYVGLRP
RSLNGAETHRLVAGMAAV	LARFGVGPGV	RVACLHGNTP	EAVLLRLAVC	WAGGCYVGLRP
RRLTGAETLEEIHRTARA	LAAOGLRPGD	GVVTLHGNGV	EAVVLRIAVO	LLGCRYAGLRP
HRITGEQLLRRCSALSRA	LADQGLRAGD	TVAYLHGNTI	DAVIARLAVÊ	ALGACYVGLRP
HRITGEQLLRRCSVLSRT	LADQGLQTGD	TVAYLHGNTI	DAVVARLAVC	ALGACYVGLRP
HRVSGEQVLRRCSALAGA	LADQGLGRGS	SVAYLHGNSI	ESVAVRLAVC	ALGACYVGLRP
RRITAGEFRDLVHRLAHA	LRAQGLERGQ	SVTLLSGNLP	EALAVRYAAN	LLGCQVNHLYN
RRITAGOFRDLVHRLAOA	LRAQGLORGO	TVTLLSGNLP	EALATRYAAN	LLGCQVN.LYD
RRVTAGEFRALVHRMARA	MRGQGIDRGR	TVTLLSGNLP	ETLAARYAAN	LIGCRVNHLYN
RRVTAGEFRALVHRMARA	MRKQGIDRGR	TVTLLSGNLP	ETVAARYAAN	LIGCRVNHLYN
RRVTAGELRDLVHRMARA	LRARGVERGQ	VVTLLSGNLP	ETIAARYAAN	LIGCAVNHLYN
RRLNAGDLRDLVYRMARA	LLTQGLGRGS	TVTLLSGNLP	ETLAARYAAN	LIGCTINHLYN
RRIGATEFRSLVYRLARA	LKERGVGPGR	GVTLLCGNLP	ETIAVRYAVN	LL <mark>G</mark> AHCNHLYN
RRLSAEEFRSLVYRLARA	LRERGVGRGK	TVTLLCGNLP	ETMAARYAVN	LLGAHFNHLYN
RRLSAEEFRSLVYRLARA	LRERGVGRGK	TVTLLCGNLP	ETMAARYAVN	LLGAHFNHLYN
RRWGAEEFRSLVHRLARA	LRERGVGRGR	TVTLLCGNLP	ETIAARYAVN	LLGAPFNHLYN
STISHGEFGELVHRIAHT	LSDHGIRSGS	TVALLSANRP	EILATRYAAN	LL <mark>G</mark> ARIVYLYE
RRITAGELVDSVHRLAHA	LDGR.VEPGR	VVALLAGNTP	EAVVARLAVI	LLGAGIAQPHE
RRITAGELVDSVHRLAHA	LDGR.VEPGR	VVALLAGNTP	EAVVARLAVI	LLGAGIAQPHE
RRITAGELVDSVHRLAHT	LDGR.VESGR	VVALLAGNTP	GAVVARLAVI	LLGAGIAQPHE
RRLTAGGIADAVHRLAHT	LDDR.VGPGD	VVALLAGNSP	EALVARYAVN	LLGAGITQPHE
VAITAGDFAAATYRLAHE	LIARGATRGT	TVTLLTGNTP	EALIARFAAG	LAGARVVNLYD
AEITAGEFAAAGYRLAHE	LIARGAARGT	TVTLLTGNTP	EALSARYAAG	LAGARVVNLYD
VGITAGEFAAATYRLAQE	LVARGVTVGS	TVTLMTGNTP	EALSARYAAG	LAGARVVNLYD
SVLTYAQVRDEVSRYAQA	YASVGLGVGS	PVAMLSGNRH	EVLIAQSANI	ITGVRFAALHP
SALTYAQVRDEVSRYAQA	YASVGLGVGS	PVAMLSGNRH	EVLIAQSANI	ITGVRFAALHP
STLTYAQVRDEVSRYAQA	YASFGLGVGS	PVAMLSGNRH	EVLIAQSTNI	ITGVRFAALHP
SILTAGALAARISQYAQV	YTAQGVRPGS	GVAMLSTNRP	EVLYSMGAYM	<b>V</b> T <b>G</b> GRNTALHP
GSLTYGELDRRSDALANA	FLEFGLEPEE	RVAMILPNRL	EAPIVDIATY	KTGAARLPINP
ETLTYADLETDSARLANA	LLECGLEPED	RVGILLSNRP	EYTVADIALA	RAGMAKVPLND

AGL76720.1_Marinactinospora_thermotolerans
WP_051713392.1_Actinoalloteichus_cyanogriseus
WP_073481158.1_Streptoalloteichus_hindustanus
WP_210842583.1_Nocardiopsis_spB62
PWV44261.1_Nocardiopsis_spL17-MgMaSL7
WP_178378553.1_Nocardiopsis_flavescens
WP_179079716.1_Streptomyces_rectiverticillatus
WP_208608527.1_Amycolatopsis_pretoriensis
WP_150246336.1_Streptomyces_albofaciens
WP_053802994.1_Streptomyces_rimosus
WP_106677009.1_Streptosporangium_nondiastaticum
WP_184939538.1_Kitasatospora_kifunensis
WP_150485975.1_Streptomyces_nitrosporeus
WP_204453063.1_Actinokineospora_baliensis
WP_121391275.1_Actinokineospora_cianjurensis
WP_075137134.1_Actinophytocola_xinjiangensis
WP_025359162.1_Kutzneria_albida
WP_083474567.1_Pseudonocardia_spAL041005-10
WP_208819870.1_Pseudonocardia_alni
WP_083658735.1_Pseudonocardia_autotrophica
WP_068798175.1_Pseudonocardia_spHH130630-07
<pre>WP_043911969.1_Kitasatospora_griseola</pre>
WP_117489816.1_Kitasatospora_xanthocidica
OKJ05654.1_Kitasatospora_spCB01950
WP_068523008.1_Tsukamurella_tyrosinosolvens
WP_126196109.1_Tsukamurella_paurometabola
WP_068744906.1_Tsukamurella_pseudospumae
WP_006543910.1_Frankia_spEUN1f
WP_161991522.1_Natronorubrum_aibiense
WP_007190185.1_Haloarcula_californiae

	80		:	٥ö			1	LO	ò					1	10						12	ò										
	TA	AVTE		AIA	AAI	SF	AI	LVI	FE	ΡS	. 1	VΕ	AR	A	A D	L	E	RV	s.	V	ΡV	V I	S	ΓG	P	гs	R.					
	MF	SPTVI	AAE	CLA	HAA	AP E	VI	L V I	ΗD	QQ	. 1	RΕ	ΕR	A	ΤE	L1	R	RA	s.	v	ΡR	VI	S	LG	ΓP.	SΑ	ь.		• •			
	VF	ATREI	K <mark>a</mark> n i	ΓLΑ	EAF	A Z	A F	V	ΥQ	ΡD	. 1	ΑN	DE	A	ΑE	L1	R	ΕV	Ρ.	. Т	ΡR	V I	S	ΓG	βP	ΑP	L.		•			
	FF	STAEI	K <mark>M</mark> K I	FLV	EGF	RP 7	A L F	ΓL]	FD	ΡD	. 1	RC	DE	Α	VS	L S	SA (	QA	G.	I	SR	ΑI	S	ΙG	V	ΑD	F.		•			
	FF	STAEI	KLKI	FLV	EGF	RP 7	↓L F	ΓL]	FD	ΡD	• •	QС	DE	Α	ΙS	L	5A	QA	G.	. I	SR	ΑI	S	ΙG	V	A D	F.		•			
	FF	ATGEI	KIKI	ΓLV	EGF	۲P ٦	ΓLΕ	L	FD	ΡD	. (	ΩS	DE	A	ΑL	L	S	RS	G.	. I	SR	ΑI	S	VG	A	5 D	F.					
3	ΚL	SADS	2 A A C	IVR	DVF	ΓF	٦A	ιI	VD	ΡA	. 3	ΥA	ΕR	A	ΑE	V.	ΓA (	QA	.Р.	. V	ΡH	V I	л	ΓG	;I:	3 T	ь.					
	KL	SADS	Q <mark>A</mark> A :	IVR	DVF	ΓF	R A I	ΙI	VD	ΡA	• 1	ΥA	QR	A.	ΑE	Ι.	[A]	ΕI	Р.	. V	ΤН	VI	JT.	ΓG	5.0	ΞE	Q.		• •			
	ΚL	SADV	2AT:	ΙVΚ	DVE	ΤZ	\AI	ΙI	VD	ΡR	.1	FD	DR	ŁA.	ΑE	L.	ΓA:	LΑ	.Р.	v	ΑT	VI	JT.	ΓG	SP.	ΑK	s.	•	• •			
	ΚL	SADV	2 <mark>A</mark> T:	T V I	DVE	ΤF	\AI	ΙI	VD	ΡR	. 1	FD	DR	A	ΑE	L.	ΓA:	LΑ	.Р.	. V	ΡT	VΙ	JT.	ΓG	βP.	ΑK	т.	•	• •			
ım	ΚL	SADV	2AT:	IMR	DVE	ΓF	RA I	L I	AD	ΡA	• 1	ΥA	GR	A.	ΑE	ЦZ	ΑE:	LΤ	G	. V	QD	V1	L	ΓG	SP(	ΞS	Ε.		• •			
	КL	SVEV	2 <mark>A</mark> A:	A V 1	DVE	ΤZ	A I	ΙI	VD	ΡR	• 1	LΑ	ΕR	Ł٨	GΕ	L	A	LΤ	P	. V	ΡT	V1	JT.	ΓG	A	ΑT	ь.	•	• •		•	
	RL	GADV	2 a k :	A V 1	DVE	TY	(A I	יע	VD	ΡR	• 1	LΑ	AR	۲A	ΤQ	I.	ΓE.	гv	Р.	. V	ΕR	V 1	V	ΓĢ	P	JΡ	ΑĽ	) A	• •		Ρ.	ΓG
	RL	GPDV	2 <b>a</b> ki	ЬMA	DVE	ΤY	ΊAΙ	ιI	VD	ΡR	• 1	LΑ	DR	۱A.	ΑT	Γ.	/ A.	ΑV	Р.	. V	ΤH	VI	v	LG	P.	A D	DI	A	• •		•	• •
	RL	G P D V (	2 A K I	AIA	DVE	Τì	(AI	ιI	VD	ΡS	• 1	LΑ	DR	Α.	ΑI	Γ,	/ A.	AV	Р.	. V	ΤН	V1	v	ΓG	B.	ΓG	G١	D'D	• •	•••	•	• •
	RL	GPDV	2 A K I	ммт	DVE	Τŀ	HT I	, I	VD	ΡT	• !	LΥ	DR	A	VQ	Ч	[S]	Т٧	А.	. <u>v</u>	DQ	F 1	т	LG	P	ΓE	VF	Ρ	• •	•••	•	•••
	GL	AVEV	2SA	IVS	DVL	TI		- V	AD	NR	• 1	LΑ	AR	Α.	ΑE	ΙV	A	EV	Ρ.	v	ΚA	LI	T.	ΓĢ	ΞP.	AS	Ξ.	•	• •	•••	•	•••
	GL	SAAA	2AR	I VD	DVE	AI		7 L.	AD	AD	• 4	AA	AA	A	ΕP	V 1	- A .	LV	Р.	G	VD	V I	W	LG	SP 1	HG	т.	•	• •	•••	•	• •
	GL	SAAA	2AR	I VD	DVE	AL		/ 14	AD	AD	• 4	AA	AA	A	ΕP	L V	- A .	LV	Ρ.	G	VD	VI	W	LG	5P	1G	т.	•	• •	•••	•	• •
	GL	SAAA	JAR.		DVE	AL	L.	<u>'</u>	AD	AD	• 4	AA	TI	A	ΕP	V 1	A.	LV	Ρ.	G	VD	V 1	W	ЬĢ	i P I	1G	т.	•	• •	•••	•	• •
	GL	SAAA	2AR		DVL	A	L.	/ <del>_</del>		AA	• 5	2A	ЕQ	A	E A	¥ 1	111	HA	Α.	. P	V Q	V 1	G	ьe	, <u>1</u> .		G.		• •	• •		:
	GM	SAPII		IAI	SVL			14		AE	- 1	Κĭ	AD	A	2 12	÷	P.	ĻĻ	D .	· •	PT		T.	LG	P	3P	DG	, D	• •		GA	A V A T
	GM	S I P V I			SVL					AE	- 1	K I	AĽ			÷	- P .		. D	· •	PI		1			3 P	GG	20	• •	. P	A I	ч. L. т. т.
	MC	SAPVI	UAE.	TVS	DAC			L 14		AB	÷÷	K I	AG	A		÷,			Ľ.	Ť	РI	Ť,	1			3 P T F	T	9.0	• •	•••	G.	гь
	MC			TME	DAC	7 T T		1.		PR	т,	T E	V P P D			+ -			E C	7 -	K T		10				÷	•	• •	•	•	• •
	MG	SIDU		(7ME	DAC			1			т.,	V E									КT	1 I T 1	20			TG	÷.	•	• •	· •	•	•••
	TG				DAU			1		P F D C		I D F T	V P F D		н D л л				E C		K D	1 1	.0			20	v	•	• •	· •	•	•••
	DL	SVTE	TDV	TLS	DAF			777		L D V C	1	ε I PT.	DD	17	ra A	T.	7 E I		Р I D T		DV	27	τΔ.	с С л т		ъD	Ď E	т	• •	•	•	· .
	ML	SATD	IEY	MLD	NSI	TS	AI	I	VG	PA	.i	FT	ET	I	ES	v	API	NV	PI	Ľ	DR	v :	S	ID	A	ΡV	SE	γĸ	LP	K R	ТН	. Э
				_			_	_						-					-	-		-		-	_							

:	130	140	150	160	170	180
AGL76720.1_Marinactinospora_thermotolerans	GRDILA.ASV	/PEGT <mark>PL</mark> RYR.	EHPEGIA	V V A F T S <mark>G T T G</mark> T	PKGVAHSSTAN	ASACVD
WP_051713392.1_Actinoalloteichus_cyanogriseus	GEDLVALAAA	AVADAPALP.	AALDRPA:	S L A Y T S <mark>G T T G</mark> S	PKGVVHGTTAI	MAACLD
WP_073481158.1_Streptoalloteichus_hindustanus	GEDLVALAGA	AQSAE <mark>PV</mark> EFT.	ADERAAT	A V G F T G <mark>G T T G</mark> R	AKGVCRAPFDI	LEACLD
WP_210842583.1_Nocardiopsis_spB62	GENLLEIAAC	KPDA <mark>PM</mark> ECV.	HRDDALSI	M V T F <mark>S S </mark> G S T G E	PKGIAHTFAS	SAGFLD
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	GENLLEIAAQ	)KPDV <mark>PM</mark> ECV.	HRDDALSI	MVTFSS <mark>G</mark> STGE	PKGIAHTFAS	SAGFLD
<pre>WP_178378553.1_Nocardiopsis_flavescens</pre>	GENLLALADO	GRPDA <mark>PL</mark> DCA.	. DRGDDVSI	MVTFTS <mark>GSTG</mark> E	PKGITHTFAS	SSGFFD
<pre>WP_179079716.1_Streptomyces_rectiverticillatus</pre>	GGDLLELAAT	ESTE <mark>PI</mark> SGL.	ARPEDIC:	S I R H T G <mark>G T T G</mark> H	IP <mark>K</mark> GICTSFER <i>I</i>	AGRMRP
WP_208608527.1_Amycolatopsis_pretoriensis	GTDLLELAAA	APTERISGL.	ARPGDIC:	S I R H T G <mark>G</mark> T <b>T G</b> H	IP <mark>K</mark> GICTDFER <i>I</i>	AGRMRP
WP_150246336.1_Streptomyces_albofaciens	GTDLLELASA	AESDE <mark>PV</mark> PSG.	ARPDDIC	I I R H T G <mark>G</mark> T <b>T G</b> H	IP <mark>K</mark> GICTTYGQV	VRTFVDDED
WP_053802994.1_Streptomyces_rimosus	GTDLLELASA	AESDE <mark>PV</mark> PSG.	ARPEDIC	I I R H T G <mark>G T T G</mark> H	IP <mark>K</mark> GICTTYGQV	VRTFVADDDED
WP_106677009.1_Streptosporangium_nondiastaticum	GTDLLELAAC	GQPAE <mark>PL</mark> SGL.	ARPEDICI	MIRHTG <mark>GTTG</mark> H	IP <mark>K</mark> AICTAFERI	LPGFQDDVD
WP_184939538.1_Kitasatospora_kifunensis	GKDLLELAAQ	)ESAE <mark>PL</mark> AGL.	ARAEDIA	I I R H T G <mark>G T T G</mark> H	I P K G I C S T F G Q (	GAAFRNGLQ
WP_150485975.1_Streptomyces_nitrosporeus	GEDLLALAAF	PQPET <mark>PV</mark> ESQ.	ARPEDLCI	MIRHSG <mark>GTTG</mark> H	I P <mark>K </mark> G V R S T F A R I	MRNTGL
WP_204453063.1_Actinokineospora_baliensis	GEDLLALAAG	GLSDE <mark>PV</mark> ASA.	ARPGDLCI	F I R H S G <mark>G T T G</mark> H	IP <mark>K</mark> GIRSTFARI	MRGLDR
WP_121391275.1_Actinokineospora_cianjurensis	GEDLLAAAA	PHSDL <mark>PV</mark> EST.	ARPGDLCI	F I R H S G <mark>G T T G</mark> H	I P <mark>K </mark> G I R S T F G R I	MRNFHR
WP_075137134.1_Actinophytocola_xinjiangensis	AEDLLALAAF	HQPDD <mark>PV</mark> TSQ.	ARPDDLCI	FIRHSG <mark>GTTG</mark> H	IP KGIRSTFARI	ALGLDQ
WP_025359162.1_Kutzneria_albida	GSDLLALAAA	A D S S <mark>P V</mark> P P H P	PVAPGDTW	S V R H T G <mark>G T T G</mark> H	IPKGILMPSGG	YRAMLE
WP_083474567.1_Pseudonocardia_spAL041005-10	DTDVTAEAAI	CASTE <mark>PV</mark> RGR.	ARPDAIE	Q I R H T G <mark>G T T G</mark> H	I P <mark>K G I T Y T F</mark> D H I	HRRGAA
WP_208819870.1_Pseudonocardia_alni	DTDVTAEAAI	CASTE <mark>PV</mark> RGR.	ARPDAIE	QIRHTG <mark>GTTG</mark> H	I P K G I T Y T F D H I	HRRGAA
WP_083658735.1_Pseudonocardia_autotrophica	DTDVTAEAA	CASTE <mark>PV</mark> RGR.	ARPDAIE	QIRHTG <mark>GTTG</mark> H	IP K G I T Y T F D H I	HRRGAA
WP_068798175.1_Pseudonocardia_spHH130630-07	RTDVLAAAEH	HASTA <mark>PI</mark> TGR.	ARLSDVE	QIRHTGGTTGH	I P K G I T Y T F G H I	HVRAAA
WP_043911969.1_Kitasatospora_griseola	GEDVIAAAAH	IRPAEPPDVR.	IRPEDDL:	SIRHTGGTTGI	PKGILTLHGP	YRAIFD
WP_117489816.1_Kitasatospora_xanthocidica	GEDVIAASAF	IRPAE <mark>P</mark> PDVR.	.VGPEDDL:	SIRHTGGTTGI	PKGILSLHGP	YRRIFD
OKJ05654.1_Kitasatospora_spCB01950	GEDVIAASAF	HRPAGPPDVR.	IGPDDDI:	SIRHTGGTTGI	PKGILTRHGT	YRLIFD
WP_068523008.1_Tsukamurella_tyrosinosolvens	GVDLTAVAAG	GFEPG <mark>PL</mark> AAH1	TADPSSTA	S L A Y TGGTTGK	SKGVMLSYSGO	GAALLR
WP_126196109.1_Tsukamurella_paurometabola	GTDLTAVAA	FEPGPLTAH1	TADPSSTA	S L A Y T G G T T G K	SKGIMLSYSG	GAALLR
WP_068744906.1_Tsukamurella_pseudospumae	GVDLTEVAAG	FEPEPLTAHI	ADPSSTA:	SLAYTGGTTGK	SKGVMLSYTGO	JASLLR
WP_006543910.1_Frankia_spEUN1f	GADLVALAAG	SYEPRPLTVPF	REDPEAIA	AIAYTGGTTGQ	PKGVMSTYRG	SAAMTQ
WP_161991522.1_Natronorubrum_aibiense	WRSAARLEDO	STAADSPSVS.	VEPDAIA	GHF Y TGGTTGD	PKGVCYTQSCI	LAINLL
wP_00/190185.1_Haloarcula_californiae	VDSFDQLLEJ	AEPVAPDVS.	VSKSTLA	GAL HIGGIIG	PEGVKHTQEN	LALNAL

		190	200	210	220
		190	200	210	229
AGL76720.1_Marinactinospora_thermotolerans		.RGPWRFLIE	PIPLSDLGGEL	AQCTL.ATGGT	VVLLEEF
WP_051713392.1_Actinoalloteichus_cyanogriseus	VARAMYG	. PPPWRFLV7	AIPLSDLGGEL	AQWIL.ACGGV	VVLREDL
WP_073481158.1_Streptoalloteichus_hindustanus	ASLTIFG	.EGPWRFLVC	CIPIADLGGEM	AEWTL.AAGGT	VVLREDF
WP_210842583.1_Nocardiopsis_spB62	NARHMYG	. PGPWRFLVA	AIPLSDLGGEI	VQWTL.AVGGT	VVLLDDF
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	NARHMYG	. PGPWRFLVA	AIPLSDLGGEI	VQWTL.AVGGT	VVLLDDF
WP_178378553.1_Nocardiopsis_flavescens	NALHMYG	. PGPWRFLVA	AIPLSDLGGEI	VQWTL.AGGGT	AVLMDDF
WP_179079716.1_Streptomyces_rectiverticillatus	QPGGEAA	APPRQLVS	STPLAHAAGLV	ADHLL.AEGGT	VVLLEDF
WP_208608527.1_Amycolatopsis_pretoriensis		DPPRQLVE	TTLAHAAGLI	ADHVL.AEGGT	VVLLEDF
WP_150246336.1_Streptomyces_albofaciens	PGDAKVLWERPADQ	ERTRLLVC	TTLAHAAGMM	ADVTL.REHGL	VVLLEDF
WP_053802994.1_Streptomyces_rimosus	AGDAKVLWERPADQ	EGNRLLV(	TTLAHAAGMM	ADVTL.REHGL	VVLLEDF
WP_106677009.1_Streptosporangium_nondiastaticum	EILEPEGKAD	DGMRQLV(	TTLAHAAGLL	ADVTLRTDGGT	VVLLDDF
WP_184939538.1_Kitasatospora_kifunensis	DS.EEAEGEE	KVERQLV(	TTLAHAAGHL	ADHTL.GEGGA	VVLLDDF
WP_150485975.1_Streptomyces_nitrosporeus	MRELSD	EVRRDLV(	CTPLSHAGGFL	ADTTL.ASGGT	VVLHDGF
WP_204453063.1_Actinokineospora_baliensis	VSGLFGG	AQPRDLLC	TPLSHAGGFI	ADSTL.VAGGT	VVLHYAF
WP_121391275.1_Actinokineospora_cianjurensis	VRATNDV	DEPRDLV(	TPLSHAGGFL	ADATL.AAGGT	VVLHYAF
WP_075137134.1_Actinophytocola_xinjiangensis	VSALFGG	HTPRDLVC	TPVSHAAGFI	ADATL.AAGGT	VVLHRTF
WP_025359162.1_Kutzneria_albida	HQATAQG	ADVVYLIS	S T T L A H G A G N L	ADGVF.AAGGT	IVLHERF
WP_083474567.1_Pseudonocardia_spAL041005-10	MRRGTGGFGSG	. DAGARLLVA	ATPVAHVGGSL	ADRTL.TDGGT	VVLQDDF
WP_208819870.1_Pseudonocardia_alni	MRRGTGGFGSG	. DAGARLLVA	ATPVAHVGGSL	ADRTL.TDGGT	VVLQDGF
WP_083658735.1_Pseudonocardia_autotrophica	MRRGAGGFGSG	. DAGARLLVA	ATPVAHVGGSL	ADRTL.ADGGT	VVLQDGF
WP_068798175.1_Pseudonocardia_spHH130630-07	MRRSMGSFGTDTAG	ESGPPRMLVA	ATPVAHAGGAL	GDRML.AGGGT	VVLLERF
WP_043911969.1_Kitasatospora_griseola		APRLLA	ATPLAHLAGVL	SDLVL.HHGGT	VVLQRSF
WP_117489816.1_Kitasatospora_xanthocidica	APIGSDPA	. AGAPRLLA	ATPLAHLAGIF	ADLTL .QHGGT	VVLQRSF
OKJ05654.1_Kitasatospora_spCB01950	DPVGNEE	APRLLA	ATPLAHLAGVF	SDLVL.YHGGL	VVLLRSF
WP_068523008.1_Tsukamurella_tyrosinosolvens	IQRTEWEW	. PEEIRFLV(	CAPLSHAGGAF	WNPTS.MTGGS	MVVLPRF
WP_126196109.1_Tsukamurella_paurometabola	IQRTEWEW	. PEEIRFLV(	CAPLSHAGGAF	WNPTS.MAGGS	MVVLPRF
WP_068744906.1_Tsukamurella_pseudospumae	IQRTEWEW	. PQEIRFLV(	CAPLSHAGGAF	WNPTS.MSGGS	MVVLPRF
WP_006543910.1_Frankia_spEUN1f	ILLTEWQW	. PKELRHLV(	CTPLSHAGAPF	IVPVL.HQGGS	VVLPKF
WP_161991522.1_Natronorubrum_aibiense	AHPADLGF	.SSDDTGLVA	ATPISHSAGTF	LLATL.IAGGT	VVLRRGF
WP_007190185.1_Haloarcula_californiae	AHAVELDI	.QPRVTL <b>LL</b> M	<u> 1 T P L P H S A G L I</u>	MAGGL.AQGSH	HVITQGF

	230	240	250	260	270	280
AGL76720.1_Marinactinospora_thermotolerans	QPDAVLEAIERI	ERATHVFLA	PNWLYQLAEHP	ALPRSDLSSLF	RVVYGGAPAV	PSRVAAA
WP_051713392.1_Actinoalloteichus_cyanogriseus	DPLATLALLERI	ERITHLFTA	PSSLYQLAEHP	DLDHF <mark>D</mark> L <b>TAL</b> F	LAVYGGAPAV	PARTSAA
WP_073481158.1_Streptoalloteichus_hindustanus	EPADILATIGA	ERTTHVFCA	PGWVYQLAEHP	ALADA <mark>DLSSL</mark> I	QIP <mark>Y</mark> GGAPS1	PARIADA
WP_210842583.1_Nocardiopsis_spB62	RADEVARTLASI	EHITHFFGS	<b>P</b> SMVHALVHEP	GLRNA <mark>D LS</mark> D <b>LF</b>	LVAYGGAPSS	PSRTADA
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	RADEVARTLGSI	EHITHFFGS	<b>P</b> SMVHALVHEQ	GLRNA <mark>D</mark> L <mark>S</mark> D LF	LVAYGGAPSS	PSRTADA
WP_178378553.1_Nocardiopsis_flavescens	RADEAVRILGRI	ERITHFFGS	<b>P</b> SMVHALVHEP	GLRNTRLPDLF	LVAYGGAVSS	<b>PSRTADA</b>
WP_179079716.1_Streptomyces_rectiverticillatus	DPGQVLAAIERI	ERITHLFLI	PPLLYQLMDHP	HADRT <mark>D</mark> T <mark>SSLF</mark>	QVIYGGCQTS	PARIANA
WP_208608527.1_Amycolatopsis_pretoriensis	DPGQVLAAIERI	ERITHLYLI	PPLLYSLVDHP	DAVRT <mark>DLSSLF</mark>	QVNYGGCKAS	PARIADA
WP_150246336.1_Streptomyces_albofaciens	DPAEVLATIER	ERITHLFLI	PPLLYQLTDHP	DAATT <mark>D</mark> T <mark>SSLF</mark>	CLTYGGCQSS	PARIADA
WP_053802994.1_Streptomyces_rimosus	DPAEVLATIERI	EKITHLFLI	PPLLYRLTDHP	DAATT <mark>D</mark> T <mark>SSLF</mark>	CLTYGGCQSS	PARIADA
WP_106677009.1_Streptosporangium_nondiastaticum	DAAEALAAIERI	ERITHMFLI	PPLLYQLMDHP	D V D S T <mark>D T S S L F</mark>	NVTYGGCQSS	PARLASA
WP_184939538.1_Kitasatospora_kifunensis	DPAEVLATIERI	ERITRLFVI	PPLMYQLLDHP	DSKDADTSSLF	TLLYGGCLAS	<b>PARLVDA</b>
WP_150485975.1_Streptomyces_nitrosporeus	DAGEVLATIARI	EEITHVLLI	PPHLYQVLDHP	DIETTDTSSMF	RITYGSCQSS	<b>PVRIAEA</b>
WP_204453063.1_Actinokineospora_baliensis	DPGEVLATIER	ERITHAFLI	PPLLYQVLDHP	GEFDTSSLQ	RITYGGTPSA	PVRIAQA
WP_121391275.1_Actinokineospora_cianjurensis	DPGEVLATIER	ERITHVFLI	PPLLYRVLDHP	GDHDTSSLS	RITYGGTPSA	PVRIAQA
WP_075137134.1_Actinophytocola_xinjiangensis	DPGEVLATIER	ERITHLFLI	PPLLYQVLDHP	ARVDTSSLF	QLTYGGTPSS	PTRIAEA
WP_025359162.1_Kutzneria_albida	APAQVLADIERI	HRVTDLFLI	PPLLYRVLDDP	ASASTDTSSLF	RFLYFGCAAS	PTRIGQA
WP_083474567.1_Pseudonocardia_spAL041005-10	EPGAYLAAVERI	ERITHGFLI	PPLIYRLLDHP	DLDRTDLSSLF	SLVYGGAPAN	PRRIAEA
WP_208819870.1_Pseudonocardia_alni	EPGAYLAAVERI	ERITHGFLI	PPLIYRLLDHP.	DLDRTDLSSLF	SLVYGGAPAN	PRRIAEA
WP_083658735.1_Pseudonocardia_autotrophica	EAGAFLAAVERI	SRITHTFLI	PPLIYRLLDHP	DLDRTDLSSLF	SLVYGGAPAN	PRRIAEA
WP_068798175.1_Pseudonocardia_spHH130630-07	DAGAWLAAVERI	ERITHTMLI	PPLLYRVLDHP	GLDRTDLSSLF	TLFYGGAPAN	PRRIAEA
WP_043911969.1_Kitasatospora_griseola	DPGAVLAAVERI	SRIIDLWVI	PLLYQLLDHP	AIGRIDLSSLF	QIMYGGAAAS	PIRIREA
WP_11/489816.1_Kitasatospora_xanthocidica	EPGAVLAAIERI	SHITDLWVI	PPLLHQLLDHP	AIGRIDLSSLF	ILSIGGSAAS	DTRIKEA
WD 069502000 1 mewharmenalla turnarina anlara	EPGAVLSAVERI	ERIIEFWVI VDTTDTMTU	PALLHQLLDHP	ALARIDLSSVF KEDEVDIGGIE	QIII GGGSAAS	DTDITEA
WP_066525008.1_TSukamurella_tyrosinosoivens	EPEAVLAAIEK	I K I I A I M L V V D T T A T M I V	PIMIIALLDHP.		TUENCAGAIS	DTDITEA
WP_120190109.1_ISUKamurella_paurometabola	EPEAVLAATEK		TMI VAL DUD		TUENCACAT	
WP_006744900.1_TSukamurella_pseudospumae	DACKULEATEN	UDITCUELU	PIMIIALLDHP.		TUENCOCOM	DCDIDEA
WD 161991522 1 Natronorubrum aibiense	DETDECANVET	HCVTWTFLV	PTML VPL LDCP	T. A A H D V S S L F	NUTYCAAPTC	PDPLODA
WP_101991922.1_Nacronorubrum_arbrense	DAEDALELVEOL		PTMUVPULDUI	EDETUDTOTIE	TTUYCAADME	PDPT PFC
woo/iyoios.i_haioaicula_callioiniae	DURUNUTEDA	SDIOWIENV				

AGL76720.1_Marinactinospora_thermotolerans
WP_051713392.1_Actinoalloteichus_cyanogriseus
WP_073481158.1_Streptoalloteichus_hindustanus
WP_210842583.1_Nocardiopsis_spB62
PWV44261.1_Nocardiopsis_spL17-MgMaSL7
WP_178378553.1_Nocardiopsis_flavescens
<pre>WP_179079716.1_Streptomyces_rectiverticillatus</pre>
WP_208608527.1_Amycolatopsis_pretoriensis
WP_150246336.1_Streptomyces_albofaciens
WP_053802994.1_Streptomyces_rimosus
WP_106677009.1_Streptosporangium_nondiastaticum
WP_184939538.1_Kitasatospora_kifunensis
WP_150485975.1_Streptomyces_nitrosporeus
WP_204453063.1_Actinokineospora_baliensis
WP_121391275.1_Actinokineospora_cianjurensis
WP_075137134.1_Actinophytocola_xinjiangensis
WP_025359162.1_Kutzneria_albida
WP_083474567.1_Pseudonocardia_spAL041005-10
WP_208819870.1_Pseudonocardia_alni
WP_083658735.1_Pseudonocardia_autotrophica
WP_068798175.1_Pseudonocardia_spHH130630-07
WP_043911969.1_Kitasatospora_griseola
WP_117489816.1_Kitasatospora_xanthocidica
OKJ05654.1_Kitasatospora_spCB01950
<pre>WP_068523008.1_Tsukamurella_tyrosinosolvens</pre>
WP_126196109.1_Tsukamurella_paurometabola
WP_068744906.1_Tsukamurella_pseudospumae
WP_006543910.1_Frankia_spEUN1f
WP_161991522.1_Natronorubrum_aibiense
WP_007190185.1_Haloarcula_californiae

290	300	310	320	330
RERM <mark>G</mark> .AVLMQN	YGTQEAA.FIAA	L T P D D H	ARRELLTAVGRPLP	HVEVEIRD
LRRL <mark>G</mark> .PRLMQN	YGTQETG.FVCA	LPEDH	GHPELLDHA <mark>G</mark> RLLP	GVEVEIRD
LEKLGRPLLVHC	YGSQEGG.WMTW	LSAED <mark>H</mark> VRA	A D R Y L <mark>L N S V G</mark> K A L P	GTEIAIRD
VSAL <mark>G</mark> .PVLMHN	YGMQEAG.FISY	LSQSDHIRA	A G R Y L <mark>L N S V G</mark> R P L P	DVEIAVRD
VSAL <mark>G</mark> .PVLMHN	YGMQEAG.FISY	LSQSD <mark>H</mark> IRA	A G R Y L <mark>L N S V </mark> G R P L P	DVEIAVRD
VAGL <mark>G</mark> .PVLMHN	YGMQEAG.FISF	LSRSDHIRA	A G R Y L <mark>L N S V G</mark> R P L P	GVEIAVRD
IRRF <mark>G</mark> .PVLMQG	YGQTETG.SISV	LGPKDHDL	I R P E R <mark>L R S A G</mark> K V L D	GVEVAVRD
IARL <mark>G</mark> .PVLLQG	YGQTESG.TISV	LEPADHDPI	R R P E R L R T A <mark>G R V L P</mark>	GVEVAIRD
VRRF <mark>G</mark> .PVLVQF	YGQNEAG.GISV	LPAGDHDPI	D N L E R L R S A G R I L P	NVEVAVRD
IRRF <mark>G</mark> .PVLVQF	YGQNEAG.GISV	LPAEDHDPI	D N L E R <mark>L R S A G</mark> R I L P	NVEVAVRD
VRRF <mark>G</mark> .HVLTQF	YGQNEAG.GISV	LSPEDHDPI	ENPARLR <b>TA<mark>G</mark>KAM</b> N	GVEIAIRD
LRRF <mark>G</mark> .PVLFQF	YGQNEAG.GIST	LSAEDHDPI	ERPDRLRSA <mark>G</mark> KVLD	GVEVEIRG
VRRF <mark>G</mark> .RVLQQS	YGQFEAG.AVSV	LPAGEHDPI	DRLDVLRTAGRVIP	GVEVEIRT
VRRF <mark>G</mark> .RVLQQG	YGQFEAG.AITV	LPAAEHDPI	DRLDVLRTAGRAVP	GVELEIRG
VRRF <mark>G</mark> .RVLQQG	YGQFEAG.AITI	LPAAEHDPI	D R L D L L R T A G R V V P	GVEVEIRD
ARRF <mark>G</mark> .PVLHQG	YGQFESG.PVSM	LTAAEHDPI	DRLDILRTAGRVLP	GVEVEIRD
LDRFG.PILFQI	YGQSEAG.GISM	LTEADH	2 R P E L L D S V G R P I P	GVEVAIRD
LRRIG. PVFTQF	YGQTEAG.GISV	LTAEEHI	RNPDLLGTVGRPIP	TTTVAITD
LRRIG. PVFTQF	YGQTEAG.GISV	LTAEEH	RNPDLLGTVGRPIP	TTRIAITD
LRRIG. PVFTQF	YGQTEAG.GISV	LTAEEH	RDPDLLGTVGRPIP	TTTIAITD
IKKIG. PVLLQF	YGQTEAG GVSV	TPADH	TRPELLGTVGRVVP	STEIAITG
FKVFG. PVLVGV	YGMSEAQ.MISA	LAPHEHER.	IGQGGQVPVGRPLP	GIEVAIRG
LKVFG.PVLVGL	YGMSEAQ.LISF	TPOERDR		GVEVAIRS
MKVFG.PVLVGV	YGMSEAQ.MISV	LIPPEHAR		NVEVVIRG
IDRIG. PVFFQF	YGQAECPMIISV		SKPERLASCGKPVP	WVNVALLD
TERIG. PVFFQF	YCOARCPMTISV		SKPERLASCGKPVP	WVNVALLD
CRIG. PKFFQF	YCOTE DOTUEN	DKEGUHDP:		WVIVALLD
DDIG DIITOF	YCOTRUDUL TT	DEPENDENCE P	DEBLI DEACODOL	WVRVALLD
DELG. PILIOF	YCOVETPDI ITV		SDER LLKSAGOPCL	MCEVKVVD
плыта.ильтбт	HOVIMIPDEIIV	EFROMEDPI	DOGERIDOSCOLPI I	PICEV K V V N

AGL76720.1_Marinactinospora_thermotolerans	1
WP_051713392.1_Actinoalloteichus_cyanogriseus	
WP_073481158.1_Streptoalloteichus_hindustanus	(
WP_210842583.1_Nocardiopsis_spB62	i
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	I
WP_178378553.1_Nocardiopsis_flavescens	i
WP_179079716.1_Streptomyces_rectiverticillatus	i
WP_208608527.1_Amycolatopsis_pretoriensis	1
WP_150246336.1_Streptomyces_albofaciens	I
WP_053802994.1_Streptomyces_rimosus	I
WP_106677009.1_Streptosporangium_nondiastaticum	H
WP_184939538.1_Kitasatospora_kifunensis	1
WP_150485975.1_Streptomyces_nitrosporeus	]
WP_204453063.1_Actinokineospora_baliensis	i
WP_121391275.1_Actinokineospora_cianjurensis	1
WP_075137134.1_Actinophytocola_xinjiangensis	1
WP_025359162.1_Kutzneria_albida	1
WP_083474567.1_Pseudonocardia_spAL041005-10	i
WP_208819870.1_Pseudonocardia_alni	2
WP_083658735.1_Pseudonocardia_autotrophica	i
WP_068798175.1_Pseudonocardia_spHH130630-07	]
WP_043911969.1_Kitasatospora_griseola	i
WP_117489816.1_Kitasatospora_xanthocidica	1
OKJ05654.1_Kitasatospora_spCB01950	1
WP_068523008.1_Tsukamurella_tyrosinosolvens	I
WP_126196109.1_Tsukamurella_paurometabola	]
WP_068744906.1_Tsukamurella_pseudospumae	]
WP_006543910.1_Frankia_spEUN1f	]
WP_161991522.1_Natronorubrum_aibiense	]
WP_007190185.1_Haloarcula_californiae	1

340	350	360	370	380	390
DSGGTLPRGAV	EVWVRSPMTMS	GYWRDPER	TAQVLS.GGWLR	GDVGTFDED	GHLHLT
TSGAVVPPGVV	EIWVRSPMTMS	GYHAD <mark>P</mark> AR	TGEVLV.DGWVR	GDLGHLSEA	DLLRIL
QDGADLPVGTV	EVCVRSTMLMR	GYWRLPEL	TAKTVR.DGWLH	GDLGRLDTE	GYLYLV
PEGGELPPGSI	EVCVRSTTLMA	GYWQRPDL	TAEVLR.EGWFL7	GDLGRVDDE	GYLYLV
PEGVELPPGSI	EVCVRSTTLMA	GYWQRPDL	TAEVLR.EGWFL	GDLGRVDDE	GYLYLV
AEGGELPAGEI	EVCVRSATIMA	GYRHRPDL	TAQVLR.EGWFR	GDLGRIDDE	GYLYLV
ADGHDLPPGEH	EICVRSGHIMQ	GYWKQPEL	TAEVLR.DGWLH1	GDVGFLDDE	GYLTVI
AADNDLPHGEHG	EICVRSGHIMQ	GYWKEPEL	TAEVLR.DGWLH1	GD IGFLDAE	GYLTIT
ENGRDLPRGEH	GEICVRSASIMQ	GYWKQPEL	TAEVLR.DGWLH1	GDIGFLDDD:	DYLTIV
ENGRDLPRGEH	EICIRSASIMQ	GYWKQPEL	TAEVLR.DGWLH1	GDIGFLDAD	DYLTIV
ESGRDLPTGEH	EVCVRSEHIMT	GYWKQPEL	TAEVLR.DGWLH1	IGD IGFLDDE	GYLTIV
EDGRALPVGEL	EICVRANGVMQ	GYWKQPEL	TAEVLR.DGWLH1	<b>'GD</b> LGHFDED	GYLTVV
ESGTVLPHGEVG	EICVRGPQVMD	GYWKDPEL	TAEVLEPDGWLHI	GDLGRLDAE	GYLTVV
ADGAALPVGQVG	SE I CV RASEVME	SYWKDPEL	TAQVLR.DGWVH	GDLGRLDEQ	GYLTIV
PDGTALPPGRVG	EICVRAGEVME	SYWKNPEL	TAQVVR.DGWVH1	GDLGRFDDH	GYLTVV
PDDAPVPAGQVG	EICVRGGGMME	SYWKNPEL	TAEVMR.GGWMH1	GDLGRLDDH	GYLTVV
EQGNDLPVGQT	EICVRSAMTSA	SYWRNPEL	TAQVWR.DGWVH1	GDAGHTDEŐ	GYLRLT
ADGRPVAPGGR	ELWVHTGMEMD	SYWKQPEL	TAATVV DGWVRI	GDVARQDAD	GYLHLL
ADGRPVVPGGR	ELWVHIGMEMD	SYWKOPEL	TAAIVV.DGWVRI	GDVARQDAD	GYLHLL
ADGRPVAPGGR	ELWVHIGMEMD	SYWKOPEL	TAATVV.DGWVR	GDVARQDAD	SILHLL
PDGARAGTGVRG	ELWVRTGAEMD	SYWKOPEL	TAETIV DGWVRI	GDVAWLDDD	GYLHLV
ADGADLPIGLQC	C THI DOD NUMA	SIWKKPDL SVWKRDDI	TASVER.DGWENT	GDVGIFDED	CVIVIC
IDCKENDTCEO	E THI DODATMA	SIWKKPDL ZVWKODEI	TAAVER DOWLN	GDVGIFDED	
DVCPPVFPCAP	FICURCRIVMO	ST NKDEO	TARVIQ.DGWLNI		CVIVIV
DACEPVERCAP	E I CVRCPI VMO	VI NKPEO	TARATA HONLH	CDVCREDDE	CVIVIV
DAGSTVEFGAF	FICVRGPLVMO	VINKPEO	TARATE HOWLH	GDIGREDDE	CVIVIV
DOGEPVPDGEP	EICVOGPLVMK	YWNKPEO	TAOALA CONLHT	GDIARVDEH	SEVITV
DIGEDLGIDEV	ELLVRAPYAFD	YFELPEA	TDETLY DOWNR	CDVGRIDDD	SYLVI.I.
ESGETVPPGOP	ETLARAAYSME	YYEMPEK	TAETVE DOWTH	GDIGKTDED	GYVYTT

AGL76720.1_Marinactinospora_thermotolerans	Ì
WP_051713392.1_Actinoalloteichus_cyanogriseus	1
WP_073481158.1_Streptoalloteichus_hindustanus	1
WP_210842583.1_Nocardiopsis_spB62	1
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	1
WP_178378553.1_Nocardiopsis_flavescens	(
WP_179079716.1_Streptomyces_rectiverticillatus	1
WP 208608527.1 Amycolatopsis pretoriensis	1
WP 150246336.1 Streptomyces albofaciens	1
WP 053802994.1 Streptomyces rimosus	1
WP 106677009.1 Streptosporangium nondiastaticum	i
WP 184939538.1 Kitasatospora kifunensis	i
WP 150485975.1 Streptomyces nitrosporeus	i
WP 204453063.1 Actinokineospora baliensis	i
WP 121391275.1 Actinokineospora cianjurensis	i
WP 075137134.1 Actinophytocola xinjiangensis	ĥ
WP 025359162.1 Kutzneria albida	3
WP 083474567 1 Pseudonocardia sp $AI.041005-10$	1
WP 208819870 1 Pseudonocardia alni	1
WP 083658735 1 Pseudonocardia autotrophica	1
WP 068798175 1 Pseudonocardia sp HH130630-07	1
WP 043911969 1 Kitagatognora grigoola	1
WP_043911909.1_Kitasatospora_giiseoia	1
OKIO5654 1 Kitasatospora sp. CP01950	1
MD 069523009 1 Mayhamunalla tunasinasaluana	1
WP_000022000.1_TSukamurella_tyrosinosoivens	1
WP_126196109.1_TSUKamurella_paurometabola	4
WP_068744906.1_Tsukamurella_pseudospumae	1
WP_006543910.1_Frank1a_spEUN1f	1
WP_161991522.1_Natronorubrum_aibiense	1
WP_00/190185.1_Haloarcula_californiae	ł

400	410	420	430	440
DRLODIIIVEAYN	VYS RR VEHVLTE		VPDPDSGBAV	CAAVVVADGA
DRAKDLIIVEAYN	VYSERVEDVLCH	HPKVLOAGVVG		CAAVVLASGT
DRAKDVIIVEAYN	VYSOEVEHVLTG	HPDVRYAAVVG	VPDHDTTEAV	YAAVVPAEGVG
DRVKDIIIVDAYN	VYPOOVEOVLTA	HPGVAQAVVVG	LPDSDTGESV	YAAVVPSTEVPAD
DRVKDIIIVDAYN	<b>V</b> YPQQVEQVLTA	HPGVAQAVVG	LPDSDTGESV	YAAV <mark>V</mark> PSTEVPAD
GRIKDIIIVDAYN	<b>V</b> YPQQV <mark>E</mark> QVLSA	HPGVAQAAVVG	VPDPDTGESV	Y A A V V P S S E I P A G
DRIKDMIVVVGGH	VYTTELEDLLNS	HPQVRQSAVFG	VRDNDRMEQV	HAAV <mark>V</mark> RAPGA
DRIKDMIVVVGGH	VYTTELEDLLNS	HPHVLHSAVFG	VRDNDHAEQV	HAAV <mark>V</mark> LSPGA
DRIKDMIVVVGGH	VYTTELEDLLNS	HPDVLQSAVFG	VRDEDRMEQV	YAAV <mark>V</mark> RAPGA
DRIKDMIVVVGGH	VYTTELEDLLNS	HPGVLQSAVFG	V R D E D R M <mark>E</mark> Q V	YATV <mark>V</mark> LADGA
DRIKDMIVVVGGH	VY T T E L E D L L N T	HPEVLQSAVFG	V P D A N R M E Q V	YAAV <mark>V</mark> RAPGS
DRLKDMIIVVGGH	VYTSELEDALNS	HPQVLQSAVFG	VPDADRMEQV.	HAVVVLAPGA
DRIKDMIAVVGGH	VYTTELEEFLHT	HPDVRHSAVYG	V R D Q G R I E R V	HATVVLAPGS
DRLKDMIVVVGGH	VYTTELEDFLTT	HPDVRAAAVYG	VPDDGGVERI.	HATLYLTPTA
DRLKDMIVVVGGH	VYTTELEDFLTT	HPDVRAAAVYG	VPDDGGVERI.	HATVVLAPDT
DRVRDMIVVVGGH	VYTAELEEFLST	HPEVRDCAVYG	VRGVDGVERI	FASVVRPGS
SRLKDMIIVVGGH	VIPVEVEEALLG	HPGIAEAAVIG	VRAEDAAELV	HAAV PRPGH
DRVKDMIVVVGGH	VIISELEDLLME WYTCELEDIIME	HPAVRHAAVF G	VPDDDGIEIV.	HAAVVIDSPA
DRVKDMIVVVGGH	VYTSELEDLIME	HPAVPHAAVEG	VPDDDGTETV	HAAVVIDSPA
DRVKDMYTVVGGH	VITSELEDILME	HPCVPOANEC	VPDCDCTEVV	HAAVVIDSFA
DRAKDLTTVVGGH	VYDSEVEELLLT	HPTVARCAVEC	VPDDOOAFHV	HVAVVDADCO
DRVKDLTTVVGGH	VYPSEVEDLLLT	HPAVARCAVEG	VRDDOEARHV	HVAVVPAPGO
DRVKDLTTVVGGH	VYPSEVEDLLVT	HPTVARCAVEG	VRDDOETEHV	HVAVVPAAGR
DRKKDMVVTGGFN	VFPREVEDVIST	HPSVASVAVIG	VPDDKWGEAV	KAVVVPRAGA
DRKKDMVVTGGFN	VFPREVEDVIST	HPSVASVAVIG	VPDEKWGEAV	KAVVVPRAGA
DRKKDMVVTGGFN	VFPREVEDVIST	HPSVASVAVIG	VPDDKWGEAV	KAVV <mark>V</mark> PRAGA
DRKKDMIVSGGFN	VFPREIEDVLST	HPAVAAAAVIG	VPDDRWGEAV	RAVV <mark>V</mark> LRDGA
DRRSDVVVTGGMN	<b>V</b> Y T G E V E H A L G E	HGAVGDVAVIG	VPHEDWG <mark>E</mark> AV	HAIV <mark>V</mark> PAERR
DRDSDVIVSGGMN	VYSVAVEEVVQQ	HDQVANVAVIG	VPDDEWG <mark>E</mark> AV	KAVVVPESD

	450	460	470	480	490
AGL76720.1_Marinactinospora_thermotolerans	DPDPEHLRA	ALVRDHLGDLH	VPRRVEFVRSI	PVTPAGK	PDKVKVRTWFTD
WP_051713392.1_Actinoalloteichus_cyanogriseus	ADADELR	EHVEAQLGEPH	VPRLIRFVDTI	PLTDGGK	PDKSALRALFAPDTPSR
WP_073481158.1_Streptoalloteichus_hindustanus	EIDVDELRA	A L V R T T L G P V H	EPKHLDVVDTI	P T T P R <mark>G K</mark>	PDKSALRTRWRAAQRA
WP_210842583.1_Nocardiopsis_spB62	RAVAFSAELAA	AVRETLGSFH	EPTRIDLLASL	PLTPRGK	P <mark>D K T A I R</mark> GQ A A D R A G S P H V
PWV44261.1_Nocardiopsis_spL17-MgMaSL7	RTVAFSAELAA	AVRETLGTVH	EPTRIDLLASM	PLTPRGK	P <mark>D K T A I R</mark> GQ A V D R A G S P H I
WP_178378553.1_Nocardiopsis_flavescens	RTAEFTAGLAA	AVRDTLGSVH	EPARIDLLASL	PLTPRGK	P <mark>D R E A L R</mark> GRAAE TAA S P R T
WP_179079716.1_Streptomyces_rectiverticillatus	KVGA <mark>DELR</mark> A	AMVREQRGAMY	EPTHITFAEQL	PLTDAGK	P <mark>D K K R L R</mark> S R T A P T Q E R S Q G
WP_208608527.1_Amycolatopsis_pretoriensis	EVDPDALRA	A <mark>MV</mark> RDQR <mark>G</mark> AMY	VPTHITFVDRL	PLTDTGK	PDKKLLRSHAAQAEGAR
WP_150246336.1_Streptomyces_albofaciens	DVDA <mark>DQLR</mark>	AMVREARGAMY	EPAHIEFVEQL	PLTDAGK	PDKKQLRAMATRTAV
WP_053802994.1_Streptomyces_rimosus	DVDA <mark>DQLR</mark>	AMVREARGAMY	EPAHIEFVEQL	PLTDAGK	PDKKQLRAQATRSAA
WP_106677009.1_Streptosporangium_nondiastaticum	ELDEKELR	EFVGNERGAMY	EPAHVSFIDAL	PLTDAGK	PDKKFLRREAEAAAATR
WP_184939538.1_Kitasatospora_kifunensis	TVEEQELRA	AMVREERGAMY	EPARITFADAL	PLTDVGK	PDKKLLRQWVEQAAV
WP_150485975.1_Streptomyces_nitrosporeus	TTTAEDLRA	ALVRDRRGELY	VPDLIEFVEAL	PLTDAGK	PDKKELRRWAA
WP_204453063.1_Actinokineospora_baliensis	TVTAESLR:	ſMVRDGRGDMY	VPDHVHLVDAL	PLTDVGK	P <mark>D</mark> KKEL <mark>R</mark> RRAADRSTAR.G
WP_121391275.1_Actinokineospora_cianjurensis	TTTPETLR:	[MVRDGRGDMY	VPDHIHLVDSL	PLTDVGK	PDKRTLRRQATSTP
WP_075137134.1_Actinophytocola_xinjiangensis	AVTEEDLR(	0 M V S T G R G A M Y	VPDRVEFVAAL	PLTDVGK	PDKKVLRGRAVG
WP_025359162.1_Kutzneria_albida	TLTEQEVR	EHVASQLGALY	TPHHVEFLDRI	PLTDAGK	PDKKVLRARH
WP_083474567.1_Pseudonocardia_spAL041005-10	PSPDELTO	GLVADRAGAMY	VPDTVEFVERI	PLTGIGK	T <mark>DKKRLR</mark> AELVGQPR
WP_208819870.1_Pseudonocardia_alni	PSPDELTO	GLVADRAGAMY	V P D T V E F V E R I	PLTGIGK	TDKKRLRAELVGQPR
<pre>WP_083658735.1_Pseudonocardia_autotrophica</pre>	PSPDELTO	GLVADRAGAMY	VPDTVQFVERI	PLTGIGK	TDKKRLRAELVGQPR
WP_068798175.1_Pseudonocardia_spHH130630-07	PTCDELGO	GLVADRAGAMY	VPRSISFVAEL	PLTGIGK	TDKKRLRAELTA
WP_043911969.1_Kitasatospora_griseola	RPELDELRA	AFVTERKGRIF	APEALHLVDEI	PLTPVGK	PDKKRLREAATR
WP_117489816.1_Kitasatospora_xanthocidica	RPDLDTLR/	AFVTERKGRIF	APEVLHLVDEI	PLTPVGK	PDKKRLRAAATA
OKJ05654.1_Kitasatospora_spCB01950	QPHLDELR/	AFVTEQKGRIF	APEALHLVDEI	PLTPVGK	PDKKRLREAVAG
WP_068523008.1_Tsukamurella_tyrosinosolvens	TVDVEELRI	DL VRDRKGAV Y	TPKTVDFVESI	PVSPLGK	PDKKALRAQYA
WP_126196109.1_Tsukamurella_paurometabola	TVDVEELRI	DL VRDRKGAV Y	TPKTVDLVESI	PVSPLGK	PDKKALRAQYA
WP_068744906.1_Tsukamurella_pseudospumae	TVDVEELRI	EL V R D R K G A V Y	TPKTVDLVDSI	PVSPLGK	PDKKALRAIYA
WP_006543910.1_Frankia_spEUNIf	SVAAEELAA	ALVKEHKGSHH	APKGVDFVDAL	PLIAVGK	PDKKALRAKHWSNTDRRVG
WP_161991522.1_Natronorubrum_aibiense		AFAGERLASYK	KPKSVDVFETL VDKGIDVUETL	PTTPYGK	IDREALRERYWADEERRIN
WP_00/190185.1_Haioarcula_californiae	TVDRANIRS	SFCRDRLSDYE	VEKSIDVVETL	PITPYGK	LDKKKLKEPYWQDQSRQIN

Figure S4. Alignment used for the phylogenetic tree construction. Alignment illustration made by using ESPript v 3.0, Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". Nucl. Acids Res. 42(W1), W320-W324 - doi: 10.1093/nar/gku316 (freely accessible online).



Figure S5. Alignment of McbA and ancestors A1-A4, derived from the phylogenetic tree in fig S1. Alignment illustration made by ESPript v 3.0, Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". Nucl. Acids Res. 42(W1), W320-W324 - doi: 10.1093/nar/gku316 (freely accessible online).



Figure S6. <sup>1</sup>H-NMR of synthesised 3-Hydroxy-N-(3-phenylpropyl)benzamide (standard).



Figure S7. <sup>13</sup>C-NMR of synthesised 3-Hydroxy-N-(3-phenylpropyl)benzamide (standard).



Figure S8. <sup>1</sup>H-NMR of the enzymatic synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide (crude NMR). As a control, <sup>1</sup>H-NMR of 3-phenyl-1-propylamine and 3-hydroxybenzoic acid are included.



Figure S9. <sup>13</sup>C-NMR of the enzymatic synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide (crude NMR). As a control, <sup>13</sup>C-NMR of 3-phenyl-1-propylamine and 3-hydroxybenzoic acid are included.



Figure S10  $^{13}$ C-NMR (left) and  $^{1}$ H-NMR (right) of 3-hydroxybenzoic acid, 3-phenyl-1-propylamine, and the enzymatic synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide (crude NMR). In the left chromatogram, a shift of the carbonyl carbon (G'), the  $\alpha$ -carbon (C') and the aromatic carbon (A') is consistent with the formation of an amide. In the right chromatogram the peak at around 8.4 ppm (S') shows the presence of an amide bond.



Figure S11. <sup>1</sup>H-NMR of synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide (standard).



Figure S12. <sup>13</sup>C-NMR of synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide (standard).



 $\label{eq:Figure S13. $^1$H-NMR of biocatalytic synthesised $3-acetyl-N-(4-(hydroxymethyl)phenethyl) benzamide} \\$ 



Figure S14. <sup>13</sup>C-NMR of biocatalytic synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide



Figure S15. DEPT spectra of biocatalytic synthesised 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide



Figure S16. <sup>1</sup>H-NMR of synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide (standard).



Figure S17. <sup>13</sup>C-NMR of synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide (standard).



Figure S18 <sup>1</sup>H-NMR of biocatalytic synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide



 $\label{eq:sigma} Figure \ {\tt S19}\ {\tt ^{13}C-NMR}\ of\ biocatalytic\ synthesised\ {\tt N-(2-(piperidin-1-ylsulfonyl)benzyl)benzyl)} benzamide$ 



Figure S20. DEPT spectra of biocatalytic synthesised N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide





Figure S21. 305 nm UV absorption calibration curve of different 3-hydroxybenzoic acid concentrations. In the equation, y is the 305 nm peak area of the acid, and x is the mM of the acid in the HPLC sample.



Figure S22. HPLC-MS (positive ionization mode, 305 nm) chromatogram of a sample taken from the upscaled biocatalytic synthesis of 3-Hydroxy-N-(3-phenylpropyl)benzamide 60 h after the start. At 0.587 min is 3-hydroxybenzoic acid, 0.748 min is 3-phenylpropylamine, and at 2.145 min is the 3-Hydroxy-N-(3-phenylpropyl)benzamide, with its expected mass of 256 m/z. For reference, all mass peaks throughout the elution are shown.



Figure S23. HPLC-MS (positive ionization mode, 305 nm) chromatogram of 3-Hydroxy-N-(3-phenylpropyl)benzamide standard. At 2.117 min is the amide, with its expected mass of 256 *m/z*. For reference, all mass peaks throughout the elution are shown.



Figure S24. 305 nm UV absorption calibration curve of different 3-acetylbenzoic acid concentrations. In the equation, y is the 305 nm peak area of the acid, and x is the mM of the acid in the HPLC sample.



Figure S25. HPLC-MS (positive ionization mode, 305 nm) chromatogram of a sample taken from the upscaled biocatalytic synthesis of 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide 40 h after the start. At 1.704 min is 3-acetylbenzoic acid, and at 2.035 min is 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide, with its expected mass of 298 *m/z*. For reference, all mass peaks throughout the elution are shown.



Figure S26. HPLC-MS (positive ionization mode, 305 nm) chromatogram of 3-acetyl-N-(4-(hydroxymethyl)phenethyl)benzamide standard. At 2.071 min is the amide, with its expected mass of 298 m/z. For reference, all mass peaks throughout the elution are shown.



Figure S27. 305 nm UV absorption calibration curve of different benzoic acid concentrations. In the equation, y is the 305 nm peak area of the acid, and x is the mM of the acid in the HPLC sample.



Figure S28. HPLC-MS (positive ionization mode, 305 nm) chromatogram of a taken from the upscaled biocatalytic synthesis of N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide 60 h after the start. At 0.640 min is benzoic acid, and at 2.341 min is N-(2-(piperidin-1-ylsulfonyl)benz



Figure S29. HPLC-MS (positive ionization mode, 305 nm) chromatogram of N-(2-(piperidin-1-ylsulfonyl)benzyl)benzamide standard. At 2.360 min is the amide, with its expected mass of 359 m/z. For reference, all mass peaks throughout the elution are shown.

Table S6. Derivation of USEtox input parameters' mean and standard deviation from in-silico model prediction and uncertainty (described by the reported log RMSE).

USEtox 2.0 input	Description	Prediction model	Mean value	Standard deviation	Comment
pKa.gain [-]	Acid dissociation constant (acid reaction)	OPERA pKa_a [-]	рКа_а	2.0	
pKa.loss [-]	Acid dissociation constant (basic reaction)	OPERA pKa_b [-]	pKa_b	1.7	
Kow [-]	Octanol-water partition ratio	OPERA LogP [-]	$10^{LogP}$	0.69	
Pvap25 [Pa]	Vapor pressure at 25C	OPERA VP [log(mm/Hg)]	$133.3 \cdot 10^{LogVP}$	1.08	with 133.3 Pa/mm.Hg <sup>-1</sup>
Sol25 [mg/L]	Water solubility at 25C	OPERA WS [log(mol/L)]	$MW \cdot 10^{LogWS} \cdot 1000$	0.81	MW = molecular weight
kdegA [1/s]	Degradation rate in air	OPERA LogOH [log(cm <sup>3</sup> /mol.s)]	$\frac{1.5e^6}{2} \cdot 10^{LogOH}$	1.14	with 1.5e <sup>6</sup> mol OH/cm <sup>3</sup>
kdegW [1/s]	Degradation rate in water	OPERA BioHL [log(d)]	$\frac{\ln(2)}{10^{BioHL} \cdot 86400}$	0.26	
kdegSd [1/s]	Degradation rate in sediment	OPERA BioHL [log(d)]	$\frac{\ln(2)}{9\cdot 10^{BioHL}\cdot 86400}$	1.5 · 0.26	with water-to-sediment extrapolation factor = 9 <sup>a)</sup> , with uncertainty increased by a factor 1.5 for intermedia extrapolation
kdegSl [1/s]	Degradation rate in soil	OPERA BioHL [log(d)]	$\frac{\ln(2)}{2\cdot 10^{BioHL}\cdot 86400}$	1.5 · 0.26	with water-to-soil extrapolation factor = 2 <sup>a)</sup> , with uncertainty increased by a factor 1.5 for intermedia extrapolation
LogHC50 [log(mg/L)]	50% hazard concentration for freshwater species	ECOSAR ChV Fish, ChV Daphnid, ChV Green Algae [mg/L]	$\frac{1}{n} \sum_{i=1}^{n} \log_{10} \frac{ChV_i}{0.4}$ i = [1: fish, 2: daphnid, 3: algae]	1.4 <sup>b)</sup>	with EC50 = ChV/0.4 <sup>c)</sup> assuming ChV~NOEC
ED50 non- cancer, ingestion [kg/lifetime]	50% chronic non- cancer effect dose via ingestion	CTV RfD NOAEL [mg/kg.d]	4.1 · 9 · 10 <sup>RfD NOAEL</sup> · 70 · 70 · 365 · 1e <sup>−6</sup>	0.9	with interspecies human-to-rat extrapolation factor = 4.1 <sup>d)</sup> , NOAEL-to-EC50 extrapolation factor = 9 <sup>d)</sup> , human bodyweight = 70 kg, human lifetime = 70 years, 1 year = 365 days, 1 mg = 1e <sup>-6</sup> kg
ED50 non- cancer, inhalation [kg/lifetime]	50% chronic non- cancer effect dose via inhalation	CTV RfD NOAEL [mg/kg.d]	ED50 <sup>non-cancer</sup>	1.5 · 0.9	with uncertainty increased by a factor 1.5 for 1:1 extrapolation from ingestion to inhalation
ED50 cancer, ingestion [kg/lifetime]	50% chronic cancer effect dose via ingestion	CTV OSF [risk per mg/kg.d]	$0.8 \cdot \frac{1}{OSF} \cdot 70 \cdot 70 \cdot 365 \cdot 1e^{-6}$	1.2	with 1/q*-to-ED50 conversion factor = 0.8 <sup>d)</sup> , human bodyweight = 70 kg, human lifetime = 70 years, 1 year = 365 days, 1 mg = 1e <sup>-6</sup> kg
ED50 cancer, inhalation [kg/lifetime]	50% chronic cancer effect dose via inhalation	CTV IUR [risk per mg/m³]	$0.8 \cdot \frac{1}{IUR} \cdot \\13 \cdot 70 \cdot 365 \cdot 1e^{-6}$	1.23	with 1/q*-to-ED50 conversion factor = 0.8 <sup>d)</sup> , inhalation rate = 13 m <sup>3</sup> /d, human lifetime = 70 years, 1 year = 365 days, 1 mg = 1e <sup>-6</sup> kg
BAF fish [-]	Bioaccumulation factor in fish	OPERA BCF [-]	$10^{LogBCF}$	0.55	

a) US EPA EPISuite (2012), b) Ping Hou et al., "Estimate Ecotoxicity Characterization Factors for Chemicals in Life Cycle Assessment Using Machine Learning Models," Environment International 135 (2020): 105393-, c) Nicolò Aurisano et al., "Extrapolation Factors for Characterizing Freshwater Ecotoxicity Effects," Environmental Toxicology and Chemistry 38, no. 11 (2019): 2568–82., d) Peter Fantke (Ed.) et al., "USEtox® 2.0 Documentation (Version 1.1)", http://usetox.org.

Table S7. Overview of uncertainty factors applied onto the model-specific uncertainty (described by the log RMSE) based on assessing the model applicability for each chemical.

Model suite	Applicability domain	Uncertainty factors
OPERA	Local applicability domain based on structural	Inside domain (AD≥0.6): 1
	similarity of query chemical to five nearest	Inside extended domain (0.4 ≤AD < 0.6): 1.5
	neighbors in training data set, ranging from 0	Outside domain (AD < 0.4): 2
	to 1	
CTV	Global applicability domain based on Z-score,	Inside domain (AD≤1): 1
	which corresponds to the number of standard	Inside extended domain (1 < AD $\leq$ 3): 1.5
	deviations	Outside domain (AD ≥ 3): 2
		Imputed value (contains metals/metalloids): 3
ECOSAR	"DomainOfApplicability" flag (no continuous	Inside domain (no AD flag): 1
	value)	Outside domain (AD flag): 2



Figure S30.USEtox characterization factors vs in-silico screening scores of safechems and subset of filtered out aromatic acids and amines A) Median freshwater ecotoxicity USEtox characterization factor of the in-silico screening environmental score of safechems and sample of 408 aromatic amines and 448 aromatic acids that were filtered out. B) Median human toxicity USEtox characterization factor of in-silico screening toxicity score of safechems and sample of 408 aromatic amines and 448 aromatic acids that were filtered out.



1





Figure S32. Toxic effects coupled to the molecular weight of amines and acids which were filtered out in the *in silico* filtering process (figure 3). The lower the value (y-axes), the more toxic. HC50 is the hazardous concentration of a chemical at which 50% of the species in an aquatic ecosystem are exposed to the chemical above their tolerance concentration, LogHC50 (log mg/L). ED50non = Human-equivalent lifetime dose per person that causes a non-cancer disease probability of 50% via either inhalation or digestion(kg/lifetime), ED50canc = Human-equivalent lifetime dose per person that causes a cancer disease probability of 50% via either inhalation or digestion (kg/lifetime).



Figure S33. Scatter plot of the USEtox CFs of amides with *in silico* hazard score equal to or lower than ten (figure 3) compared to the sum of the CFs of their constituent acid and amine moieties. The results demonstrate a clear deviation between the amide CF and the combined CFs of the acid and amine. This observation underscores the importance of directly evaluating the environmental impact of the final product, rather than solely relying on the CFs of its building blocks. The amides in the plot passed the initial filtering, and the acids and amines are their safechem building blocks.

	McbA											
	1b	2b	3b	4b	5b	6b	7b	8b	9b	10b	11b	12b
1a	23.8 ± 0.5	31.2 ± 0.9	8.0 ± 0.1	27.6 ± 0.5	0 ± 0	0 ± 0	41.7 ± 0.0	0 ± 0	29.0 ± 3.2	34.3 ± 0.2	$1.5 \pm 0.1$	34.7 ± 9.0
2a	5.0 ± 0.1	2.5 ± 0.0	$0.4 \pm 0.0$	1.5 ± 0.0	0 ± 0	0.1 ± 0.0	0.8 ± 0.1	1.5 ± 0.0	1.7 ± 0.1	2.6 ± 0.7	0.2 ± 0.0	$1.1 \pm 0.1$
3a	2 ± 0.0	$1.8 \pm 0.0$	0.2 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.6 ± 0.0	0.4 ± 0.0	0.2 ± 0.0	0 ± 0	0.2 ± 0.0
4a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
5a	0 ± 0	0.5 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	$1.4 \pm 0.0$	0 ± 0	0.2 ± 0.0	0 ± 0	0.2 ± 0.0
6a	29.9 ± 0.5	38.1 ± 1.9	7.0 ± 0.2	12.4 ± 0.3	2.3 ± 0.2	$0.4 \pm 0.1$	10.4 ± 0.3	10.2 ± 0.1	11.6 ± 0.2	6.7 ± 0.3	10.6 ± 0.3	10.4 ± 0.7
7a	1.6 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
8a	0 ± 0	4.7 ± 0.3	0.6 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	2.7 ± 0.0	1 ± 0.0	0.8 ± 0.1	0 ± 0	0.5 ± 0.0
9a	$1.4 \pm 0.1$	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
10a	15.1 ± 0.9	37.0 ± 1.0	10.1 ± 0.3	19.0 ± 0.2	1.1 ± 0.2	0.1 ± 0.0	8.2 ± 0.7	12.0 ± 0.1	13.0 ± 0.1	9.8 ± 0.3	4.8 ± 0.4	9.1 ± 0. 7
11a	7.3 ± 0.3	4.9 ± 0.2	0 ± 0	1.2 ± 0.1	0 ± 0	0 ± 0	0.3 ± 0.0	1.6 ± 0.0	$1.3 \pm 0.0$	0.5 ± 0.0	0 ± 0	0.5 ± 0.0
12a	1.5 ± 0.1	0.3 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	$0.4 \pm 0.0$	0 ± 0	0 ± 0
<b>13</b> a	4.5 ± 0.0	3.8 ± 0.1	$1.9 \pm 0.1$	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.8 ± 0.0	0.8 ± 0.0	$0.4 \pm 0.0$	0 ± 0	0 ± 0
14a	6.3 ± 0.0	4.5 ± 0.2	0.3 ± 0.0	$1.1 \pm 0.0$	0 ± 0	0 ± 0	0.6 ± 0.0	1.2 ± 0.1	$1.9 \pm 0.1$	0.8 ± 0.0	0.6 ± 0.1	0.9 ± 0.2
15a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.6 ± 0.2	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table S8. Conversion by McbA (in %). The first column refers to the acids, and the first row refers to the amines. Conversion is calculated from the amide peak area divided by total acid and amide peak area (DAD) and based on two replicates.

						A1						
	1b	2b	3b	4b	5b	6b	7b	8b	9b	10b	11b	12b
1a	21.9 ± 0.0	17.8 ± 0.2	3.3 ± 0.0	8.0 ± 0.2	0 ± 0	0.7 ± 0.0	0 ± 0	0 ± 0	11.2 ± 0.5	3.4 ± 0.1	18.3 ± 0.3	9.6 ±2. 9
2a	$1.0 \pm 0.0$	0.2 ± 0	0 ± 0	0 ± 0	0 ± 0	0.1 ± 0.0	0 ± 0	0 ± 0	0.2 ± 0.0	0 ± 0	0.9 ± 0.0	0 ± 0
3a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
4a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
5a	0 ± 0	0.2 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
6a	4.2 ± 0.0	2.4 ± 0.0	0.3 ± 0.0	0.6 ± 0.3	0.2 ± 0.0	0.7 ± 0.2	0 ± 0	0.7 ± 0.2	0 ± 0	0.2 ± 0.0	4.8 ± 0.8	0 ± 0
7a	1.5 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
8a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
9a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
10a	5.4 ± 0.0	1.9 ± 0.2	0.8 ± 0.0	0.9 ± 0.1	0 ± 0	0.2 ± 0.0	0.8 ± 0.0	1.3 ± 0.1	$1.4 \pm 0.1$	0.1 ± 0.0	3.4 ± 0.0	3.6 ± 0.1
11a	0.3 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
12a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
1 <b>3</b> a	0.5 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
14a	0.2 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	$0.1 \pm 0.0$	0 ± 0
15a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	$1.6 \pm 0.6$	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table S9. Conversion by A1 (in %). The first column refers to the acids, and the first row refers to the amines. Conversion is calculated from the amide peak area divided by total acid and amide peak area (DAD) and based on two replicates.

	Α2											
	1b	2b	3b	4b	5b	6b	7b	8b	9b	10b	11b	12b
1a	6.2 ± 0.6	1.7 ± 0.0	22.4 ± 1.2	13.4 ± 0.3	0 ± 0	0 ± 0	16.5 ± 1.1	0 ± 0	1.2 ± 0.0	35.9 ± 0.7	0.9 ± 0.0	51.5 ± 15.6
2a	0.3 ± 0.0	0 ± 0	3.2 ± 0.7	$1.2 \pm 0.1$	0 ± 0	0 ± 0	2.3 ± 0.1	5.1 ± 0.2	0.3 ± 0.0	6.5 ± 0.1	0 ± 0	14.1 ± 0.3
3a	0 ± 0	0 ± 0	0.8 ± 0.2	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.2 ± 0.0	0 ± 0	0.4 ± 0.0	0 ± 0	2.6 ± 0.2
4a	0 ± 0	0 ± 0	0.5 ± 0.0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
5a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
6a	$1.3 \pm 0.0$	0.9 ± 0.1	6.1 ± 0.0	2.9 ± 0.0	0 ± 0	0±0	7.8 ± 0.0	6.4 ± 1.0	0 ± 0	5.9 ± 0.1	0 ± 0	24.6 ± 2.3
7a	$1.3 \pm 0.0$	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
8a	0 ± 0	0 ± 0	12.2 ± 1.7	0 ± 0	0 ± 0	0 ± 0	4.8 ± 0.2	9.9 ± 0.5	0 ± 0	7.7 ± 0.0	0 ± 0	8.9 ± 0.6
9a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
10a	1.2 ± 0.0	0.2 ± 0.0	10.7 ± 0.5	3.2 ± 0.0	0.1 ± 0.0	0 ± 0	5.7 ± 0.0	4.9 ± 0.3	0.6 ± 0.0	5.9 ± 0.1	0 ± 0	9.1 ± 0.6
11a	$1.1 \pm 0.0$	0 ± 0	9.0 ± 0.6	0.9 ± 0.0	0 ± 0	0 ± 0	10.6 ± 1.4	1.5 ± 0	0.6 ± 0.0	8.0 ± 1.8	0 ± 0	4.6 ± 0.0
12a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.5 ± 0.1	0 ± 0	0 ± 0
13a	0.8 ± 0.0	0 ± 0	10.5 ± 0.9	0 ± 0	0 ± 0	0 ± 0	18.4 ± 2.2	1.9 ± 0.0	0.6 ± 0.0	6.1 ± 0.1	0 ± 0	6.5 ± 0.0
14a	0.7 ± 0.0	0 ± 0	4.1 ± 0.5	1.2 ± 0.4	0 ± 0	0 ± 0	5.0 ± 0.0	3.5 ± 0.1	1.5 ± 0.1	10.9 ± 0.1	0 ± 0	9.7 ± 0.8
<b>15</b> a	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	2.1 ± 0.1	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table S10. Conversion by A2 (in %). The first column refers to the acids, and the first row refers to the amines. Conversion is calculated from the amide peak area divided by total acid and amide peak area (DAD) and based on two replicates.

Compound	SMILES	Hazard factor	Ecohazard factor	used in experimental synthesis
Aniline	NC1=CC=CC=C1	2	0	yes
amine8634	NCc1ccccc1S(=O)(=O)N1CCCCC1	3	1	yes
amine729	NC1(c2cccc2)CCNC1	3	1	yes
amine6672	NC1CCN(S(=O)(=O)c2cccc2)C1	3	1	yes
amine6328	NCCCc1ccccc1	1	1	yes
amine516	NCCCN1Cc2cccc2C1	3	1	yes
amine3327	NCc1ccccc1CN1CCCCC1	2	1	yes
amine3289	NCCc1ccc(CCN)cc1	4	1	yes
amine2540	NCc1ccccc1CN1CCCC1CO	3	1	yes
amine2119	NCCc1ccc(CO)cc1	2	1	yes
amine14086	Cc1ccccc1CCCN	2	1	yes
amine13521	CN(CCN)Cc1ccccc1	1	1	yes
amine13034	NCCc1ccccc10	3	0	yes
amine12828	NCc1ccc(S(=O)(=O)NCC2CCCO2)cc1	2	1	yes
amine1217	COCCNC(=O)C(N)Cc1ccccc1	4	1	yes
amine11946	CN(C)S(=O)(=O)Cc1ccc(CN)cc1	3	1	yes
2-aminopyridine	NC1=CC=CC=N1	2	1	yes
amine9676	NC1(c2cccc2)CCC(=O)NC1	4	1	
amine8732	NC1CCCS(=O)(=O)c2ccccc21	3	1	
amine8721	COc1ccc(S(=O)(=O)N2CCC(N)CC2)cc1	2	1	
amine8666	NC1CCc2cc(C(=O)O)ccc2C1	4	1	
amine8369	NC1CCCN(c2ccc(C(=O)O)cc2)C1	3	1	
amine7654	CC(c1ccc(S(C)(=O)=O)cc1)N1CCC(N)CC1	3	1	
amine7539	NCCCCc1ccccc1	1	1	
amine6912	C=Cc1ccc(CCN)cc1	2	1	
amine6247	Cc1cc(C)cc(C(O)CN)c1	2	1	
amine5745	CN(CCN)S(=O)(=O)c1ccc(F)cc1	4	1	
amine5247	NCCCc1cccc(O)c1	2	1	

amine5086	COc1ccc(CN)cc1	3	1	
amine4400	COc1ccc(CCN)c(F)c1	3	1	
amine429	NCC(O)c1ccccc1	1	0	
amine4167	Cc1cccc(N2CCCC(N)C2=O)c1	2	1	
amine3884	NCCN(CC(=O)O)Cc1ccccc1	4	1	
amine3444	CC(N)c1ccc(N2CCCS2(=O)=O)cc1	3	1	
amine3280	CCN(CC)Cc1ccccc1CN	4	1	
amine3274	Cc1ccc(CCN)cc1	2	1	
amine3222	CCN(CC)Cc1ccc(CN)cc1	3	1	
amine3084	CCN(CCCN)c1ccccc1	4	1	
amine3076	CN(CCCN)Cc1ccccc1	1	1	
amine3005	COC(=O)C(N)c1cc(C)cc(C)c1	3	1	
amine2998	COC(=O)C(N)c1ccc2c(c1)CCO2	3	1	
amine1943	NC(c1ccccc1)C1CCOC1	4	1	
amine14452	NC1CCCN(C(=O)c2ccccc2)C1	4	1	
amine14127	CN(C)CCOc1ccccc1CN	4	1	
amine14103	CN(C)c1ccc(CCN)cc1	4	1	
amine14084	NCCc1cccc(O)c1	2	1	
amine13680	NCCN1Cc2cccc2C1	3	1	
amine1335	NCCc1ccccc1F	2	1	
amine13327	COc1ccc(CCCN)cc1	3	1	
amine13246	NCCc1ccc(O)cc1	2	1	
amine13174	NCCCOCCc1ccccc1	1	1	
amine12952	NC(c1ccccc1)C1CCCO1	3	1	
amine12947	CCN(CC)Cc1cccc(CN)c1	3	1	
amine12928	CN(C)CCOc1cccc(CN)c1	4	1	
amine12902	CNS(=O)(=O)c1cc(F)c(Cl)c(CN)c1	4	2	
amine12829	Nc1cccc(C(N)CC(=O)O)c1	3	1	
amine1276	CCc1ccccc1CCN	1	1	
amine11623	CC(N)(CO)c1cccc(N)c1	3	1	

amine11341	NCc1cccc2c1CCNC2	4	1	
amine10961	NCc1cccc(F)c1	2	0	
amine10593	NCC1CCc2cccc2CC1	4	1	
amine10438	NCc1cccc2c1NCCC2	4	1	

Table S12. List of safechem acids from in silico screening.

Compound	SMILES	Hazard factor	Ecohazar d factor	used in experimental synthesis			
acid9360	O=C(O)c1cccc(C(=O)O)c1	2	0	yes			
acid8305	O=C(O)c1ccccc1	1	0	yes			
acid7954	O=C(O)c1cccc(O)c1	2	0	yes			
acid6946	O=C(O)c1ccccc1O	2	0	yes			
acid25277	O=C(O)c1ccc(F)c(O)c1	2	0	yes			
acid24861	CC(=O)c1cccc(C(=O)O)c1	1	0	yes			
acid24207	Cc1cccc(C)c1C(=O)O	2	0	yes			
acid24135	O=C(O)C(=O)c1ccccc1	2	0	yes			
acid23854	O=C(O)c1cccc(B(O)O)c1	3	0	yes			
acid20978	O=C(0)c1ccccc1CO	3	0	yes			
acid18918	O=C(O)C1CC(O)CN1C(=O)c1ccc2c(c1)CCC 2	2	1	yes			
acid18600	O=C(O)c1ccc2c(c1)COC2	3	0	yes			
acid14665	CC(=O)c1ccc(C(=O)O)cc1	2	0	yes			
acid11933	CC(=O)c1ccccc1C(=O)O	2	0	yes			
acid10321	O=C(O)c1ccc(O)cc1	2	0	yes			

4-methoxybenzoic acid	COc1ccc(C(=O)O)cc3	4	0	yes
acid22013	NC1Cc2ccc(C(=O)O)cc2C1	2	1	
acid9525	O=C(O)c1c(F)cccc1F	2	0	
acid8330	O=C(O)c1ccc(C(=O)O)cc1	2	0	
acid7972	O=C(O)c1ccc2c(c1)CCNCC2	4	1	
acid7295	O=C(O)c1ccc2c(c1)CC2	3	0	
acid4638	O=C(O)c1ccc2c(c1)CNC2	4	1	
acid6413	CC(=O)C(NS(=O)(=O)c1ccc(C)c(C)c1)C(O) O	4	1	
acid7343	CCc1ccccc1CC(=O)O	3	0	
acid2256	CCc1ccccc1C(=O)O	2	0	
acid6407	CC(=O)C(NS(=O)(=O)c1ccc(C)cc1)C(O)O	4	1	
acid8776	Cc1cc(O)cc(C(O)O)c1C(O)O	4	1	
acid6280	COc1cc(C=O)ccc1C(O)O	4	0	
Nicotinic acid	O=C(O)c1cccnc3	2	0	
hydrocinnamic acid	0=C(0)CCc1ccccc3	0	0	

	1	r	r	1	Exa	ct Mass	r	1		r	r	1	r
		amine11946	amine1217	amine14086	amine2119	amine2540	amine3327	amine6328	amine6672	amine729	amine8634	aniline	Phenethyl- amine
		1b	2b	3b	4b	5b	6b	7b	8b	9b	10b	11b	12b
2-Naphthoic acid	1a	382.1	376.2	303.2	305.1	374.2	358.2	289.2	380.1	294.2	408.2	247.1	275.1
4-methoxy- benzoicacid	2a	362.1	356.2	283.2	285.1	354.2	338.2	269.1	360.1	296.2	388.1	227.1	255.1
acid10321	3a	348.1	342.2	269.1	271.1	340.2	324.2	255.1	346.1	282.1	374.1	213.1	241.1
acid11933	4a	374.1	368.2	295.2	297.1	366.2	350.2	281.1	372.1	308.2	400.1	239.1	267.1
acid14665	5a	374.1	368.2	295.2	297.1	366.2	350.2	281.1	372.1	308.2	400.1	239.1	267.1
acid18600	6a	374.1	368.2	295.2	297.1	366.2	350.2	281.1	372.1	308.2	400.1	239.1	267.1
acid18918	7a	485.2	479.2	406.2	408.2	477.3	461.3	392.2	483.2	419.2	511.2	350.2	378.2
acid23854	8a	376.1	370.2	297.2	299.1	368.2	352.2	283.1	374.1	310.1	402.1	241.1	269.1
acid24207	9a	360.2	354.2	281.2	283.2	352.2	336.2	267.2	358.1	294.2	386.2	225.1	253.1
acid24861	10a	374.1	368.2	295.2	297.1	366.2	350.2	281.1	372.1	308.2	400.1	239.1	267.1
acid25277	11a	366.1	360.1	287.1	289.1	358.2	342.2	273.1	364.1	300.1	392.1	231.1	259.1
acid6946	12a	348.1	342.2	269.1	271.1	352.3	324.2	255.1	346.1	282.1	374.1	213.1	241.1
acid7954	13a	348.1	342.2	269.1	271.1	340.2	324.2	255.1	346.1	282.1	374.1	213.1	241.1
acid8305	14a	332.1	326.2	253.1	255.1	324.2	308.2	239.1	330.1	266.1	358.1	197.1	225.1
acid9360	15a	376.1	370.2	297.1	299.1	368.2	352.2	283.1	374.1	310.1	402.1	241.1	269.1

### **Notes and references**

- 1. Steffensky M, Mühlenweg A, Wang ZX, Li SM, Heide L. Identification of the Novobiocin Biosynthetic Gene Cluster of *Streptomyces spheroides* NCIB 11891. Antimicrob Agents Chemother. 2000 May;44(5):1214–22.
- 2. Pojer F, Li SM, Heide L. Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. Microbiology (Reading). 2002 Dec;148(Pt 12):3901–11.
- 3. Galm U, Schimana J, Fiedler HP, Schmidt J, Li SM, Heide L. Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tü 6040. Archives of Microbiology. 2002 Aug 1;178(2):102–14.
- 4. Schmutz E, Steffensky M, Schmidt J, Porzel A, Li SM, Heide L. An unusual amide synthetase (CouL) from the coumermycin A1 biosynthetic gene cluster from Streptomyces rishiriensis DSM 40489. Eur J Biochem. 2003 Nov;270(22):4413–9.
- 5. Wang ZX, Li SM, Heide L. Identification of the Coumermycin A 1 Biosynthetic Gene Cluster of *Streptomyces rishiriensis* DSM 40489. Antimicrob Agents Chemother. 2000 Nov;44(11):3040–8.
- Kalan L, Gessner A, Thaker MN, Waglechner N, Zhu X, Szawiola A, et al. A Cryptic Polyene Biosynthetic Gene Cluster in Streptomyces calvus Is Expressed upon Complementation with a Functional bldA Gene. Chemistry & Biology. 2013 Oct;20(10):1214–24.
- 7. Rui Z, Petříčková K, Škanta F, Pospíšil S, Yang Y, Chen CY, et al. Biochemical and Genetic Insights into Asukamycin Biosynthesis. Journal of Biological Chemistry. 2010 Aug;285(32):24915–24.
- McAlpine JB, Bachmann BO, Piraee M, Tremblay S, Alarco AM, Zazopoulos E, et al. Microbial Genomics as a Guide to Drug Discovery and Structural Elucidation: ECO-02301, a Novel Antifungal Agent, as an Example <sup>+</sup>. J Nat Prod. 2005 Apr;68(4):493–6.
- 9. Zhang W, Bolla ML, Kahne D, Walsh CT. A Three Enzyme Pathway for 2-Amino-3-hydroxycyclopent-2-enone Formation and Incorporation in Natural Product Biosynthesis. J Am Chem Soc. 2010 May 12;132(18):6402–11.
- 10. Winn M, Rowlinson M, Wang F, Bering L, Francis D, Levy C, et al. Discovery, characterization and engineering of ligases for amide synthesis. Nature. 2021 May 20;593(7859):391–8.