

# An unusual flavin-dependent halogenase from the metagenome of the marine sponge *Theonella swinhoei* WA

Duncan R. M. Smith<sup>a, d</sup>, Agustinus R. Uria<sup>b, e</sup>, Eric J. N. Helfrich<sup>b</sup>, Daniela Milbredt<sup>c</sup>, Karl-Heinz van Pée<sup>c</sup>, Jörn Piel<sup>\*b</sup>, Rebecca J. M. Goss<sup>\*a</sup>

<sup>a</sup>School of Chemistry, University of St Andrews, St Andrews, KY16 9ST, United Kingdom

<sup>b</sup>Institute of Microbiology, Eidgenössische Technische Hochschule Zürich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland

<sup>c</sup>Technische Universität Dresden, 01062 Dresden, Germany

<sup>d</sup>currently researching at The Manchester Institute of Biotechnology(MIB) The University of Manchester, M1 7DN

<sup>e</sup>currently researching at The Research and Development Center for Marine and Fisheries Product Competitiveness and Biotechnology.

Jalan KS Tubun Petamburan VI, Jakarta 10260, Indonesia.

rjmg@st-andrews.ac.uk

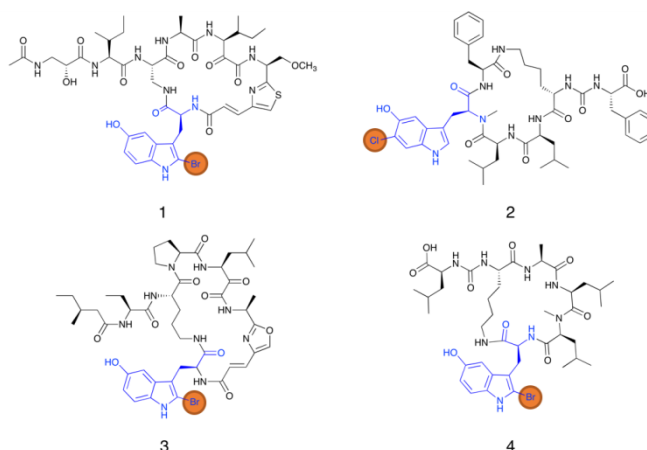
RECEIVED DATE (automatically inserted by publisher);

**Abstract:** Uncultured bacteria from sponges have been demonstrated to be responsible for the generation of many potent, bioactive natural products including halogenated metabolites.<sup>1</sup> The identification of gene clusters from the metagenomes of such bacterial communities enables the discovery of enzymes that mediate new and useful chemistries and allows insight to be gained into the biogenesis of potentially pharmacologically important natural products. Here we report a new pathway to the keramamides (krm); the first functional evidence for the existence of a distinct producer in the *Theonella swinhoei* WA chemotype is revealed, and a key enzyme on the pathway, a unique flavin dependent halogenase with a broad substrate specificity, and with potential as a useful new biocatalytic tool is described.

**Introduction:** Sponges have been shown to be a rich source of bioactive molecules, including halometabolites, and an exciting treasure trove in which to search for new enzymes. Halogenation represents an important strategy for tuning and improving bioavailability and bioactivity of molecules, and is attracting increasing interest from academia and industry alike. Over 4000 halogenated natural products have been identified to date, many of which exhibit potent bioactivities.<sup>1</sup> Keramamides and related compounds (**Figure 1**) are a unique series of halogenated marine non-ribosomal peptide natural products that have been shown to be cytotoxic against murine and human cancer cell lines.<sup>2</sup>

Due to conferring favorable properties, halogenation is used prolifically by both agrochemical and pharmaceutical industries, with around 20% of small molecule drugs and 30% of agrochemical compounds being halogenated.<sup>3</sup> In 2013, the top-revenue drug in the USA was the 2,3-dichlorophenyl-containing antipsychotic aripiprazole (Abilify).<sup>4</sup> Furthermore, the introduction of halogens into molecules can provide sites for selective functionalization through cross-coupling chemistries.<sup>5</sup> Synthetic conditions for halogenation are often harsh and non-regioselective, utilize noxious reagents, or generate undesirable by-products<sup>6</sup> and

there is an increasing demand for the discovery and development of halogenases as new tools for (bio-)catalysis. In contrast to synthetic approaches, halogenases are able to regioselectively introduce halogens under mild and green conditions without generating harmful waste. As a result, halogenases are attracting much attention.<sup>7–15</sup>



**Figure 1** Symbiont-derived natural products containing the halo-hydroxytryptophan motif (highlighted in blue) that have been identified from extracts of *T. swinhoei* chemotypes. Two distinct halo-hydroxytryptophan motifs are observed: 2-bromo-5-hydroxytryptophan and 6-chloro-5-hydroxytryptophan. Keramamide H (**1**), keramamide A (**2**), konbamide (**3**), orbiculamide A (**4**).

Keramamides contain rare halo-hydroxytryptophans motifs. Notably, the marine sponge *Theonella swinhoei* has been identified as the source of several natural products that contain this moiety, including konbamide,<sup>14</sup> and orbiculamide<sup>15</sup> (e.g., **1–4**, Figure 1) as well as the keramamides.<sup>2,16</sup> Intriguingly, distinct variants of these compounds occur in different chemotypes of the sponge. Orbiculamide A (**4**) (containing a 2-bromo-5-hydroxytryptophan moiety) is found in the chemotype *T. swinhoei* Y, which is named to reflect the characteristic

yellow color of the sponge interior (Figure 1).<sup>17</sup> *T. swinhoei* chemotype WA (white interior from Hachijo Island, Japan) was not previously reported to contain such peptidic natural products.

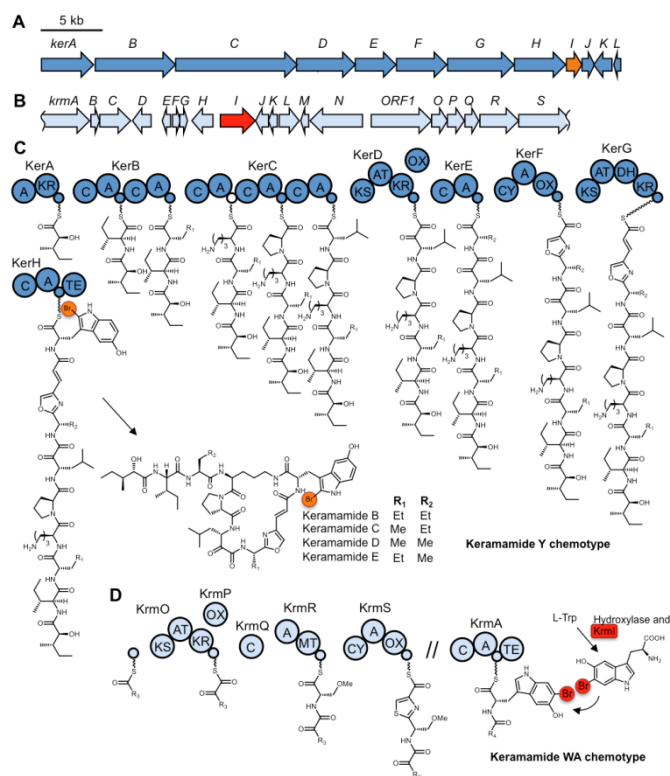
## RESULTS AND DISCUSSION

### A new *krm* gene cluster from *T. swinhoei* WA

By identifying the biosynthetic genes, it was recently demonstrated that keramamide-type peptides and many other natural products isolated from *T. swinhoei* Y are produced by an as-yet-uncultured bacterial symbiont of the sponge, *Candidatus* ‘Entotheonella factor’ TSY1.<sup>17</sup> The sponge *T. swinhoei* WA examined in the current study was also previously shown to contain filamentous ‘Entotheonella’ symbionts.<sup>17</sup> In this sponge, ‘*Entotheonella sert*’ TSWA1 is the source of the homodimeric polyketide actin inhibitor misakinolide A.<sup>18</sup> Here, using an enriched ‘Entotheonella’ cell fraction prepared from *T. swinhoei* WA, PCR analysis with primers specific for adenylation (A) domain-encoding regions of nonribosomal peptide synthetase (NRPS) genes revealed the presence of homologous genes. These were closely related to those of the keramamide (*ker*) biosynthetic genes previously identified in ‘E. factor’ associated with *T. swinhoei* Y. To isolate the corresponding genomic locus, a fosmid library was constructed from the enriched ‘Entotheonella’ cell fraction of the WA chemotype and screened, resulting in the isolation of fosmid pADU2a. Sequencing of his fosmid confirmed the presence of NRPS and polyketide synthase (PKS) genes sharing high similarity and architectural features with the *ker* cluster (Figures 2 and S1, Table S3). Notable differences of the *krm* biosynthetic gene cluster to the *ker* system, included *krmR*, and *krmI*. Keramamides identified from *T. swinhoei* WA exhibit an *O*-methylserine extension as present in keramamide H (1, Figure 1), and *krmR* is predicted to encode an NRPS module activating and incorporating *O*-methylserine. The C-terminal region of *KrmI* resembles flavin-dependent tryptophan halogenases (FDTHs). FDTHs are typically found encoded within a natural product biosynthetic pathway gene cluster, and many are known to halogenate indole, phenol, and pyrrole motifs. Though halotryptophans occur fairly frequently in natural products, such as halohydroxytryptophans are far less common.

Challenged by the low concentration of keramamides in the sponge tissue, which is itself of limited availability, and the current unculturable nature of ‘Entotheonella’ from either *T. swinhoei* WA or Y1, we set out to use comparative MS analysis to analyze the chemistry produced by the sponge microbiomes and to glean evidence for the activity of the newly identified *krm* gene cluster. Strikingly, our analysis revealed the presence of numerous keramamide congeners from each system. Figures S2 and S3 show a molecular network analysis of extracts from both chemotypes. Whilst the MS<sup>2</sup> characteristics of some *T. swinhoei* WA compounds

matched those of known keramamides, others are unique and observed for the first time.



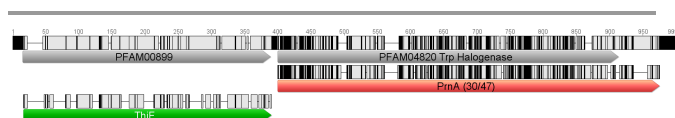
**Figure 2.** (A) The previously identified keramamide (*ker*) biosynthetic gene cluster from *T. swinhoei* Y. (B) Newly identified *krm* keramamide-type gene cluster from *T. swinhoei* WA containing the unusual 5-hydroxy-6-halotryptophan gene, (C) Biosynthetic model for the *ker* pathway. (D) Predicted function of the *krm* pathway, in which the halogenase *KrmI* functionalizes the 6 position of the indole. (See table S1 for further information on the genes identified in the *krm* pathway.<sup>14</sup>)

The detection of more than a dozen different keramamide congeners in both chemotypes might indicate an inherent flexibility of the *krm*-encoded adenylation domains, often observed for NRPS pathways, which seem to accept series of similar amino acids. From the comparative MS network analysis, the position of halogenation cannot be determined Challenged by the limited amounts of material available, as NMR analysis of the WA keramamides would not be a possible method for us to determine the regiochemistry of the halogenation, we next set out to directly investigate the unusual associated halogenase *KrmI*.

### Bioinformatic analysis of *krmI*

Bioinformatic analysis of the gene *krmI* at the translated amino acid level using BLAST<sup>19</sup> and HHPred<sup>20</sup> (SI Tables S2 and S3) indicated that the C-terminus of the protein *KrmI* is homologous to known halogenases including RebH (30% identity, 48% similarity over homologous region), PrnA (30% / 47%), and CmdE (21% / 35%), and contains the well-characterized sequence motifs for FDTHs - GxGxxG and

WxWxIP - as well as mechanistically relevant residues (Lys443 and Glu696 align to Lys79 and Glu346 from PrnA) (SI figure S3).<sup>21–23</sup> Unusually, KrmI appears to include a second, N-terminal domain-like region with distant homology to members of the ThiF enzyme family (Pfam family *ThiF* PF00899). Typically, these enzymes utilize ATP to activate carboxylic acids by adenylation.<sup>24,25</sup> Whilst sequence identity to characterized protein sequences is low, the alignment revealed that in KrmI, this domain is truncated and the majority of residues involved in substrate and co-factor binding are likely to be absent, although we must be cautious making definite claims based on alignments of such distantly related proteins. Only one other sequence displaying this unusual combination of truncated ThiF-like and FDTH domains has been deposited in the NCBI non-redundant protein sequence database; this hypothetical protein is uncharacterized and no information on the context of this gene is available.<sup>26</sup> The ThiF family is similar to the E1-like proteins that have been implicated in the specific binding of peptide substrates in ribosomal peptide natural product biosynthesis,<sup>27,28</sup> and it is possible that the C-terminal domain serves a similar function here.



**Figure 3.** A graphical representation of KrmI with its two enzymatic domains in grey (top, numbered by translated amino acid position), and homologous regions of Pfam protein families PF00899 (*ThiF*, green) and PF04820 (*PrnA* Trp Halogenase, red) that were identified using HHpred.<sup>20</sup> Similarity between KrmI domains and the PFAM representatives to which they are compared are indicated by a grey-scale similarity representation with black showing high levels of similarity and light grey showing low levels of similarity.

### Production of KrmI

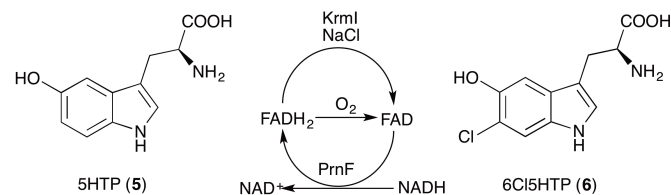
To determine whether *krmI* encodes an active halogenase, and whether both domains are necessary for activity, two expression constructs were prepared. The first encoded the full-length gene, whilst the second encoded only the C-terminal domain (truncated to the first in-frame ATG codon preceding the FDTH domain). Both constructs produced little soluble protein when expressed in *Escherichia coli* unless the chaperone proteins GroEL/S, DnaK/J and GrpE were co-expressed. Though the production of many other halogenases in heterologous expression hosts do not require the use of chaperones,<sup>29–32</sup> chaperones GroEL/S has previously been reported to improve the solubility and yield of RebH when produced in *E. coli*.<sup>11</sup>

Coexpression of genes encoding GroEL/S, DnaK/J and GrpE along with KrmI resulted in adequate levels of production of soluble KrmI (Figures S6 and S7). Attempts to produce either protein as an N-terminal maltose binding protein (MBP) fusion resulted in complete loss of solubility, whilst

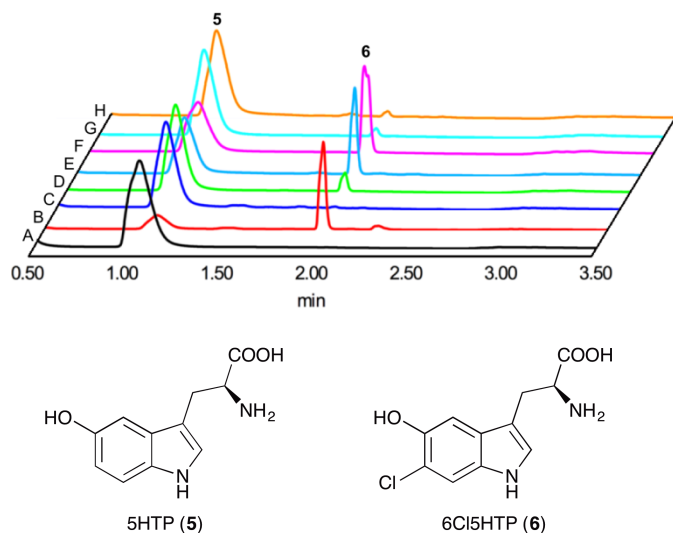
the presence of a polyhistidine tag at either terminus had only a slightly detrimental effect on solubility. A significant proportion of the soluble protein did not bind to nickel nitrilotriacetic acid (NiNTA) resin during immobilized metal affinity purification (IMAC), indicating that the polyhistidine tag may have low accessibility. IMAC purification of this enzyme is low-yielding (approximately 1 mg protein per L *E. coli* culture). We demonstrated, however, that the crude cell lysate (CCL) containing KrmI could be readily utilized as a biocatalyst, without any detrimental effect on the reaction, and therefore KrmI was used as a component of the CCL throughout the majority of the work reported here.

### Halogenation assays & product characterization

The gene was shown to encode a functional halogenase as *in vitro* assays demonstrated that the full-length protein was able to convert 5-hydroxytryptophan (5HTP, **5**) to a single new product when the enzyme was utilized as a component of a CCL (Scheme 1). LC-HRMS was used to identify the accurate mass of this product, suggesting a chemical formula of C<sub>11</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub>Cl and indicating that 5HTP (**5**) had been regioselectively chlorinated by KrmI ([M+H]<sup>+</sup> = C<sub>11</sub>H<sub>12</sub>N<sub>2</sub>O<sub>3</sub>Cl calculated: 255.0536; measured: 255.0532 Δ: +0.2 ppm). Use of the truncated protein resulted in only trace conversion to 6Cl5HTP (Figure 4C), and the purified enzymes also showed greatly reduced activity (Figure 4D and 4G) compared to the full-length protein. Hence it was decided to utilize KrmI as a component of the CCL for the rest of this study.



**Scheme 1.** KrmI converts 5HTP (**5**) to 6Cl5HTP (**6**). A simplified catalytic cycle for the required co-factor FAD/H<sub>2</sub>, which is generated *in situ* by the flavin reductase PrnF, is shown.



**Figure 4.** Halogenation of 5HTP (**5**) achieved utilizing various KrmI preparations. The conversion of 5HTP to 6Cl5HTP, as monitored by UPLC, at an absorbance at 300 nm is depicted. Assay conditions are described in the SI. (A) 5HTP (**5**) standard, (B) Full-length KrmI as a component of a CCL (0.6 mU/mL). (C) truncated C-terminal domain of KrmI as a component of the CCL (trace activity). (D) Purified full-length KrmI (20  $\mu$ M, 0.1 mU/mL). (E) Purified full-length KrmI + *E. coli* CCL (20  $\mu$ M, 0.3 mU/mL), (F) Purified full-length KrmI + catalase + SOD (20  $\mu$ M, 0.6 mU/mL). (G) Purified C-terminal domain (20  $\mu$ M, <0.1 mU/mL). (H) Purified C-terminal domain + catalase + SOD (20  $\mu$ M, <0.1 mU/mL). 1 mU is the amount of enzyme that converts 1 nmol of 5HTP (**5**) to 6Cl5HTP (**6**) per minute under assay conditions.

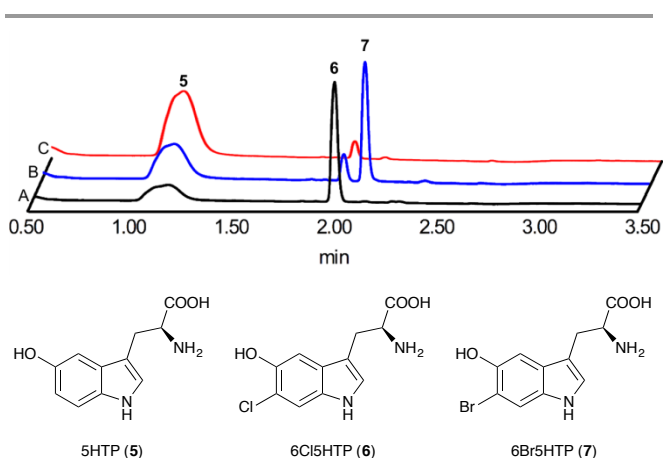
Exclusively in the case of the full-length purified protein, it was possible to restore activity by supplementing with either a CCL prepared from *E. coli* expressing an empty vector (**Figure 4E**) or with catalase and superoxide dismutase (SOD) (**Figure 4F**), implying that the protein may be harmed by reactive oxygen species (ROS) generated *in situ*. Other FDTHs reported in the literature have been active upon purification without the need for the inclusion of any further additives to counter the effect of ROS.<sup>29,30,33</sup>

In order to fully characterize the product and identify the regiochemistry of halogenation, the reaction was scaled up to 10 mg, which allowed the product to be purified and analyzed by <sup>1</sup>H NMR. As the aromatic region of the <sup>1</sup>H and COSY NMR spectra were devoid of coupling, it could be concluded that halogenation has taken place at position 6 and the three observed singlet aromatic resonances correspond to the protons at positions 2, 4, and 7 of the indole ring. 6-chloro-5-hydroxytryptophan (6Cl5HTP) has previously been described in the literature,<sup>34</sup> and the <sup>1</sup>H NMR characterization agrees with that reported here. The single product generated by the KrmI-catalyzed regioselective chlorination of 5HTP (**5**) was therefore confirmed as

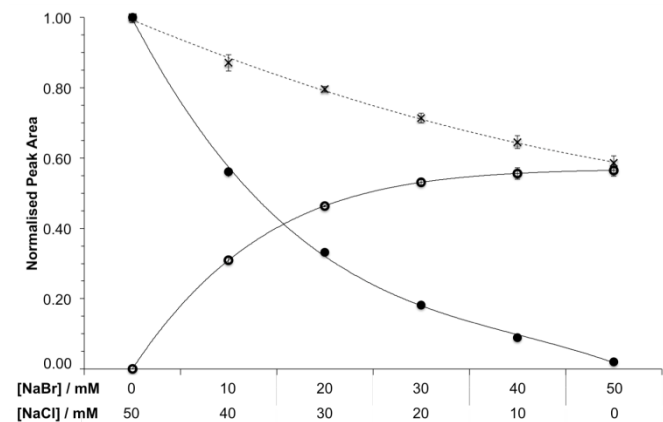
6Cl5HTP (**6**) (**Scheme 1**), a likely biosynthetic precursor to keramamide A (**2**) and other congeners.

### Halide and substrate scope

Having determined that KrmI is able to chlorinate 5HTP, we next sought to explore whether it could also catalyze bromination or iodination of the same substrate. As the enzyme was used in the form of a CCL, it was not practical to completely eliminate endogenous sodium chloride from the preparation; however, we found that the enzyme was able to tolerate the addition of salt to the CCL. Therefore, we performed assays in the presence of additional sodium chloride, sodium bromide, and potassium iodide. As is typical for FDTH,<sup>30,33,35</sup> KrmI is able to utilize chloride and bromide but not iodide, with reduced activity when incorporating bromide (**Figures 5 and 6**).



**Figure 5.** UPLC chromatograms (UV absorbance at 300 nm) monitoring the products of 5HTP (**5**) halogenation by KrmI in the presence of NaCl (A), NaBr (B), and KI (C). A small amount of NaCl is present in all assays, leading to the production of 6Cl5HTP. In the presence of bromide, an additional new product is formed, which was determined to be 6Br5HTP.

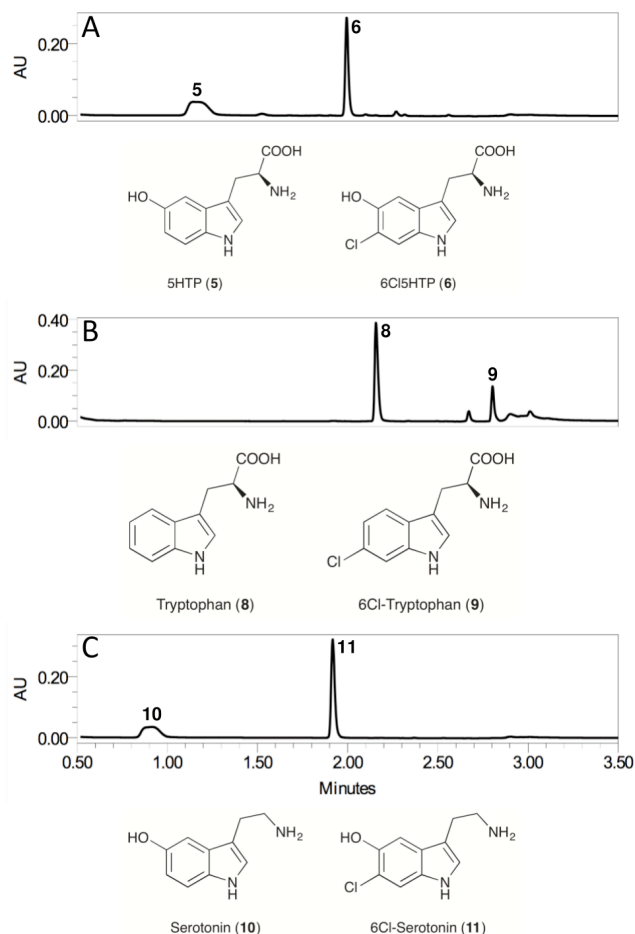


**Figure 6.** Product distribution of 5HTP (**5**) halogenation by KrmI in the presence of varying additional concentrations of Cl<sup>-</sup> and Br<sup>-</sup>. UPLC peak areas (300 nm) are plotted as normalized values. 6Cl5HTP: filled circles, 6Br5HTP: open circles, total 6-halo-5-HTP: crosses & dashed line. Error bars indicate the standard deviation of triplicate measurements and are in many cases obscured by the data markers.

The brominated product was isolated from a scaled-up reaction and characterized by  $^1\text{H}$  NMR, which indicated that regioselectivity had been retained and the product was 6-bromo-5-hydroxytryptophan (6Br5HTP). By combining sodium bromide and sodium chloride in varying ratios in the assay, we were able to determine that bromination occurred in preference to chlorination despite chloride being present in higher concentrations and a decrease in overall conversion of the starting material (**Figure 6**).

FDTH are known to halogenate a range of substrates,<sup>11,12,32</sup> typically tryptophan derivatives and other aromatic moieties. To initiate a study of the substrate scope for this unusual halogenase we first explored whether the enzyme might halogenate tryptophan and serotonin, as they are closely related to 5HTP. Indeed, KrmI was able to halogenate serotonin and 5HTP with approximately equal efficiency in terms of time and conversion levels, whilst tryptophan was less well tolerated as a substrate (**Figure 7**). To our knowledge, the only other FDTH with a preference for a substituted tryptophan is KtzR, which installs chlorine at the 6-position of 7-chlorotryptophan in the biosynthesis of kutzneride.<sup>36</sup>

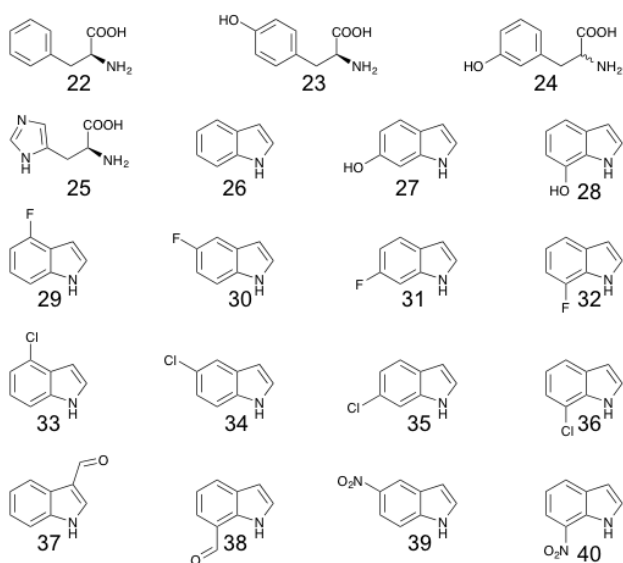
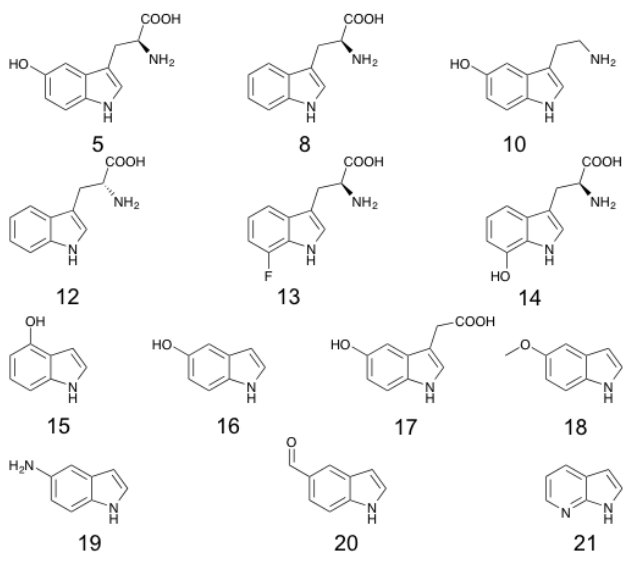
Encouraged by these initial results, we extended our study of the substrate scope of KrmI. Again we utilized crude lysate containing the full-length protein in incubations with a variety of potential substrates (**Figure 8**). We observed good levels of conversion for the substrates L-5HTP (**5**), serotonin (**10**), and 5-methoxyindole (**18**); and lower levels of conversion for L-tryptophan (**8**), D-tryptophan (**12**) and 5-aminoindole (**19**). We determined that L-7-fluorotryptophan (**13**), L-7-hydroxytryptophan (**14**), 4-hydroxyindole (**15**), 5-hydroxyindole (**16**), 5-hydroxyindole-3-acetic acid (**17**), 5-formylindole (**20**), and 7-azaindole (**21**) could also be processed by KrmI, but at trace level as indicated by LC-MS measurements. All substrates were converted to a single new mono-chlorinated product, with the exception of 5-aminoindole (**19**), which was converted into two halogenated products (LC-HRMS analysis of all halogenated substrates is included in the SI). This indicates that KrmI is able to regioselectively halogenate a range of substrates, although due to low levels of product, the position of halogenation has not been confirmed in all cases. The acceptance of 5-hydroxyindole (**16**), 5-methoxyindole (**18**) and 5-aminoindole (**19**) but not of 5-fluoroindole (**30**) or 5-fluorotryptophan (generated *in situ* from **30**) may suggest that specificity within the active site could arise from the requirement for a hydrogen bond acceptor within the ring, rather than steric requirements, as 5-fluorotryptophan is isosteric to the preferred substrate 5HTP. There is clearly considerable substrate flexibility as 5-formylindole (**20**) and 4-hydroxyindole (**15**) were also accepted. The accommodation of both L- and D-tryptophan (**8** and **12**) demonstrates that this hydrogen bonding interaction is not an absolute requirement, and that interactions with the amino acid moiety are flexible.



**Figure 7** UPLC chromatograms of the halogenation of 5HTP ((A), 300 nm), tryptophan ((B), 280 nm), and serotonin ((C), 300 nm). The identities of the labelled peaks have been confirmed by LC-HRMS and NMR.

## Conclusions

Through metabolite extraction and MS network analysis we show that *T. swinhoei* WA (white chemotype), like its yellow counterpart, *T. swinhoei* Y produces keramamides. Whilst the keramamide series of compounds produced by *T. swinhoei* WA and Y partially overlap, they also show distinct chemotype-specific characteristics. The discovery of the new biosynthetic gene cluster (e.g. *krmI*) from the metagenome of the sponge *T. swinhoei* WA explains the difference in the suites of compounds produced in each system. The *krm* genes, though showing homology to the previously identified *ker* gene cluster, exhibit distinct differences; a key difference is the presence of *krmI*. Bioinformatic analysis of *krmI* indicated it to be a flavin-dependent halogenase. Functional analysis of the encoded protein product, revealed a catalytically active halogenase with an *in vitro* substrate preference for 5-hydroxytryptophan over tryptophan, generating 6-halo-5-hydroxytryptophan as a single regioisomer, indicating that this halogenase operates to generate the 6-halo containing keramamides. It is likely that a second, as yet unidentified halogenase acts to generate the



**Figure 8.** Substrate specificity of KrmI. Substrates shown above the bold line were halogenated, whilst those shown below the bold line were not. Substrates **13** and **14** were generated *in situ* from the corresponding indole by the native *E. coli* tryptophan synthase. The substrates that were halogenated include: L-5HTP (**5**), L-tryptophan (**8**), serotonin (**10**), D-tryptophan (**12**), 7-fluorotryptophan (**13**), 7-hydroxytryptophan (**14**), 4-hydroxyindole (**15**), 5-hydroxyindole (**16**), 5-hydroxyindole-3-acetic acid (**17**), 5-methoxyindole (**18**), 5-aminoindole (**19**), 5-formylindole (**20**), and 7-azaindole (**21**). The enzyme did not process: L-phenylalanine (**22**), L-tyrosine (**23**), DL-metatyrosine (**24**), L-histidine (**25**), indole (**26**), 6-hydroxyindole (**27**), 7-hydroxyindole (**28**), 4-fluoroindole (**29**), 5-fluoroindole (**30**), 6-fluoroindole (**31**), 7-fluoroindole (**32**), 4-chloroindole (**33**), 5-chloroindole (**34**), 6-chloroindole (**35**), 7-chloroindole (**36**), 2-formylindole (**37**), 7-formylindole (**38**), 5-nitroindole (**39**) and 7-nitroindole (**40**).

2-halo containing keramamides observed within the *T. swinhoei* Y system.

KrmI is the first halogenase identified to use L-5-hydroxytryptophan as its preferred substrate. Furthermore another unique characteristic is the unusual N-terminal region of the protein. Though no function has been assigned to the N-terminal domain, in our hands, the mono-domain truncation demonstrated considerably lower activity compared to its full-length counterpart. Our initial investigations of KrmI as a potential tool for biocatalysis demonstrate that it has promisingly broad substrate specificity. KrmI can accommodate a range of indole-derived substrates. Future work will include crystallographic analysis which may provide insight into substrate flexibility and informed engineering of the enzyme to further develop it for biotechnological applications. KrmI is an efficient catalyst for the production of halogenated 5HTP and serotonin, and due to its inherent substrate flexibility presents a promising starting point for the generation of a range of chlorinated and brominated products following protein engineering.

It is notable that halogenation was almost exclusively observed at the 6-position of the indole ring in these studies, whilst in the keramamide congeners halogenation is observed at both the 2- and the 6-positions. It seems likely, therefore, that there is either another halogenase in the keramamide gene clusters that is responsible for halogenation at the 2-position of the indole ring, or KrmI possesses an activity *in vivo* which is not observed *in vitro*. It is tempting to postulate that the C-terminal domain may be responsible to binding a peptide substrate, and that the true activity of KrmI would not be observed when the enzyme acts on a free amino acid substrate rather than a peptide substrate. The instability of 2-halotryptophans, in which loss of halogen is promoted by the free carboxylate or even amine positioning over the indole ring<sup>37,38</sup> makes it likely that halogenation at the 2-position does not occur on free tryptophan substrates. Such halogenation is more likely to occur once the tryptophan substrate is integrated into a peptide, where both N- and C-termini are functionalized as amides.

This study adds further functional evidence to the role of symbiotic bacteria in the production of sponge-derived natural products. Whilst keramamides have previously been assigned to ‘E. factor’ in the chemotype *T. swinhoei* Y,<sup>17</sup> the source of the halogenated congeners in *T. swinhoei* WA remains unknown. In another study of the WA chemotype, the polyketide misakinolide A was assigned to the symbiont ‘E. sarta’ TSWA1 by single-cell genome sequencing;<sup>18</sup> however, analysis of that sequence data did not reveal the presence of keramamide genes, suggesting that these compounds are likely to be produced by either a different ‘Entotheonella’ phylotype or another symbiotic bacterium found in the sponge.

## METHODS

Isolation of *krm* genes from the metagenome of *T. swinhoei* WA

DNA was isolated from a previously described<sup>17</sup> ‘Entotheonella’-enriched bacterial fraction of *T. swinhoei* WA following a procedure adapted from Syn and Swarup. PCR using degenerate primers was used to amplify gene fragments corresponding to NRPS A domains, which were purified, ligated into pBlueScript II (SK-) (Stratagene) and sequenced. A second round of PCR using specific primers enabled identification of the fosmid pADU2a containing the *krm* cluster.

#### Metabolic analysis

Metabolites were extracted from *T. swinhoei* Y and WA tissue samples and analyzed by LC-MS-MS using a Thermo QExactive mass spectrometer operating in positive ionization mode and using data-dependent acquisition of fragmentation spectra. This data was processed using the Global Natural Products Social Molecular Networking web server ([www.gnps.ucsd.edu](http://www.gnps.ucsd.edu)). Novel keramamide analogues were assigned based on spectral similarity to known keramamide analogues.

#### Vector preparation

A PCR product containing *krmI* and appropriate restriction sites was ligated into a vector derived from pET21a using standard techniques. The sequence of vectors was confirmed by restriction digest and sequencing.

#### Protein production

*E. coli* BL21-DE3, expressing chaperone genes from the plasmid pG-KJE8 (TaKaRa Bio) was transformed with the vector encoding KrmI. An overnight culture of this was used to inoculate LB medium containing appropriate antibiotics and chaperone inducers. This was grown at 37 °C to an OD of 0.6, then cooled to 16 °C and induced by addition of IPTG. The cells were harvested by centrifugation and lysed by sonication. The crude lysate was clarified by centrifugation. The protein was purified using Ni-NTA affinity chromatography, eluting with imidazole. Proteins were analyzed by denaturing polyacrylamide gel electrophoresis according to standard methods.

#### Halogenation assays

Assays containing purified KrmI or CCL containing KrmI along with PrnF, FAD, substrate, NaCl, and NADH, (optionally also superoxide dismutase and catalase) in a final volume of 0.1 mL, were incubated at 28 °C for 90 minutes then quenched by addition of acidic methanol. Samples were analyzed by UPLC or LCMS. For preparation of purified substrates, the assays were scaled up to 10 mL and purified by semi-preparative HPLC.

#### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

Full experimental methods, supporting figures S1 to S9, supporting tables S1 to S4 and characterization of isolated products

NMR spectra

LC-HRMS chromatograms

#### Acknowledgements

We thank S Matsunaga for providing a sample of *T. swinhoei* WA; and the University of St Andrews United College scholarship (DRMS), EU BlueGenics (Seventh Framework Programme, Collaborative project “BlueGenics”, Grant no. 311848 RJMG and JP), the SNF (Grant no. 205321\_165695 to JP), the Helmut Horten Foundation (JP), and ERAIB (Grant no. 031A338A KHVP and RJMG) for funding.

#### References

- (1) Gribble, G. W. (2010) Naturally Occurring Organohalogen Compounds - A Comprehensive Update. Springer, Vienna.
- (2) Kobayashi, J., Itagaki, F., Shigemori, I., Takao, T., and Shimonishi, Y. (1995) Keramamides E, G, H, and J, new cyclic peptides containing an oxazole or a thiazole ring from a *Theonella* sponge. *Tetrahedron* 51, 2525–2532.
- (3) Herrera-Rodriguez, L. N., Khan, F., Robins, K. T., and Meyer, H.-P. (2011) Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation. *Chim. Oggi* 29, 31–33.
- (4) Top 100 Drugs for 2013 by Sales - U.S. Pharmaceutical Statistics, <https://www.drugs.com/stats/top100/sales>, (accessed July 2015).
- (5) Mahoney, K. P. P., Smith, D. R. M., Bogosyan, E. J. A., and Goss, R. J. M. (2014) Access to High Value Natural and Unnatural Products through Hyphenating Chemical Synthesis and Biosynthesis. *Synthesis* 46, 2122–2132.
- (6) Eissen, M., and Lenoir, D. (2008) Electrophilic Bromination of Alkenes: Environmental, Health and Safety Aspects of New Alternative Methods. *Chem. – Eur. J.* 14, 9830–9841.
- (7) Grüşchow, S., Smith, D. R. M., Gkotsi, D. S., and Goss, R. J. M. (2014) Halogenases, in *Biocatalysis in Organic Synthesis 3* (Faber, K., Fessner, W.-D., and Turner, N. J., Eds.), pp 313–359. Thieme, Stuttgart.
- (8) Smith, D. R. M., Grüşchow, S., and Goss, R. J. M. (2013) Scope and Potential of Halogenases in Biosynthetic Applications. *Curr. Opin. Chem. Biol.* 17, 276–283.
- (9) Brown, S., and O’Connor, S. E. (2015) Halogenase Engineering for the Generation of New Natural Product Analogues. *ChemBioChem* 16, 2129–2135.
- (10) Timmins, A., and de Visser, S. P. (2015) Enzymatic Halogenases and Haloperoxidases: Computational Studies on Mechanism and Function. *Adv. Protein Chem. Struct. Biol.* 100, 113–151.
- (11) Payne, J. T., Andorfer, M. C., and Lewis, J. C. (2013) Regioselective Arene Halogenation using the FAD-

- Dependent Halogenase RebH. *Angew. Chem. Int. Ed.* *52*, 5271–5274.
- (12) Frese, M., Guzowska, P. H., Voß, H., and Sewald, N. (2014) Regioselective Enzymatic Halogenation of Substituted Tryptophan Derivatives using the FAD-Dependent Halogenase RebH. *ChemCatChem* *6*, 1270–1276.
- (13) Frese, M., and Sewald, N. (2015) Enzymatic Halogenation of Tryptophan on a Gram Scale. *Angew. Chem. Int. Ed.* *54*, 298–301.
- (14) Kobayashi, J., Ichi, Sato, M., Murayama, T., Ishibashi, M., Wälchi, M. R., Kanai, M., Shoji, J., and Ohizumi, Y. (1991) Konbamide, a Novel Peptide with Calmodulin Antagonistic Activity from the Okinawan Marine Sponge *Theonella* sp. *J. Chem. Soc. Chem. Commun.* 1050–1052.
- (15) Fusetani, N., Sugawara, T., Matsunaga, S., and Hirota, H. (1991) Orbiculamide A: A Novel Cytotoxic Cyclic Peptide from a Marine Sponge *Theonella* sp. *J. Am. Chem. Soc.* *113*, 7811–7812.
- (16) Kobayashi, J., Sato, M., Ishibashi, M., Shigemori, H., Nakamura, T., and Ohizumi, Y. (1991) Keramamide A, A Novel Peptide from the Okinawan Marine Sponge *Theonella* sp. *J. Chem. Soc. [Perkin 1]* 2609–2611.
- (17) Wilson, M. C., Mori, T., Rückert, C., Uria, A. R., Helf, M. J., Takada, K., Gernert, C., Steffens, U. A. E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E. J. N., Brachmann, A. O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., and Piel, J. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* *506*, 58–62.
- (18) Ueoka, R., Uria, A. R., Reiter, S., Mori, T., Karbaum, P., Peters, E. E., Helfrich, E. J. N., Morinaka, B. I., Gugger, M., Takeyama, H., Matsunaga, S., and Piel, J. (2015) Metabolic and Evolutionary Origin of Actin-Binding Polyketides from Diverse Organisms. *Nat. Chem. Biol.* *11*, 705–712.
- (19) Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* *215*, 403–410.
- (20) Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res.* *33*, W244–W248.
- (21) Zhu, X., De Laurentis, W., Leang, K., Herrmann, J., Ihlefeld, K., van Pée, K.-H., and Naismith, J. H. (2009) Structural Insights into Regioselectivity in the Enzymatic Chlorination of Tryptophan. *J. Mol. Biol.* *391*, 74–85.
- (22) Flecks, S., Patallo, E. P., Zhu, X., Ernyei, A. J., Seifert, G., Alexander, Dong, C., Naismith, J. H., and van Pée, K.-H. (2008) New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed. Engl.* *47*, 9533–9536.
- (23) Dong, C., Flecks, S., Unversucht, S., Haupt, C., Pée, K.-H. van, and Naismith, J. H. (2005) Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science* *309*, 2216–2219.
- (24) Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H.-J., Costello, C. A., Backstrom, A. D., McLafferty, F. W., and Begley, T. P. (1998) Thiamin Biosynthesis in *Escherichia coli*. *J. Biol. Chem.* *273*, 16555–16560.
- (25) Duda, D. M., Walden, H., Sfondouris, J., and Schulman, B. A. (2005) Structural Analysis of *Escherichia coli* ThiF. *J. Mol. Biol.* *349*, 774–786.
- (26) Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., Marsac, N. T. de, Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K. W., Han, C. S., Rubin, E. M., Eisen, J. A., Woyke, T., Gugger, M., and Kerfeld, C. A. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci.* *110*, 1053–1058.
- (27) Burkhart, B. J., Hudson, G. A., Dunbar, K. L., and Mitchell, D. A. (2015) A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nat. Chem. Biol.* *11*, 564–570.
- (28) Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W. E., Jaspars, M., and Naismith, J. H. (2015) Structural Analysis of Leader Peptide Binding Enables Leader-free Cyanobactin Processing. *Nat. Chem. Biol.* *11*, 558–563.
- (29) Keller, S., Wage, T., Hohaus, K., Hölzer, M., Eichhorn, E., and van Pée, K.-H. (2000) Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chem. Int. Ed.* *39*, 2300–2302.
- (30) Zehner, S., Kotsch, A., Bister, B., Süßmuth, R. D., Méndez, C., Salas, J. A., and van Pée, K.-H. (2005) A Regioselective Tryptophan 5-Halogenase Is Involved in Pyrroindomycin Biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* *12*, 445–452.
- (31) Bitto, E., Huang, Y., Bingman, C. A., Singh, S., Thorson, J. S., and Phillips, G. N. (2008) The Structure of Flavin-Dependent Tryptophan 7-halogenase RebH. *Proteins Struct. Funct. Bioinforma.* *70*, 289–293.
- (32) Shepherd, S. A., Karthikeyan, C., Latham, J., Struck, A.-W., Thompson, M. L., Menon, B. R. K., Styles, M. Q., Levy, C., Leys, D., and Micklefield, J. (2015) Extending the Biocatalytic Scope of Regiocomplementary Flavin-Dependent Halogenase Enzymes. *Chem. Sci.* *6*, 3454–3460.
- (33) Yeh, E., Garneau, S., and Walsh, C. T. (2005) Robust in vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 3960–3965.
- (34) Zhang, P., Liu, R., and Cook, J. M. (1995) An Enantiospecific synthesis of L(-)- and D(+)-6-Chloro-5-Hydroxytryptophan: an Unusual Amino acid Residue from the Cyclic Hexapeptide Keramamide a. *Tetrahedron Lett.* *36*, 7411–7414.
- (35) Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N. L., and Walsh, C. T. (2005) Dichlorination of a Pyrrolyl-S-carrier Protein by FADH<sub>2</sub>-dependent Halogenase PhtA during Pyoluteorin Biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 13843–13848.
- (36) Heemstra, J. R., and Walsh, C. T. (2008) Tandem Action of the O<sub>2</sub>- and FADH<sub>2</sub>-Dependent Halogenases KtzQ and KtzR Produce 6,7-Dichlorotryptophan for Kutzneride Assembly. *J. Am. Chem. Soc.* *130*, 14024–14025.



- (37) Phillips, R. S., and Cohen, L. A. (1983) Synthesis of 2-Bromo-L-tryptophan and 2-Chloro-L-tryptophan. *Tetrahedron Lett.* 24, 5555–5558.
- (38) Hinman, R. L., and Bauman, C. P. (1964) Reactions of N-Bromosuccinimide and Indoles. A Simple Synthesis of 3-Bromooxindoles. *J. Org. Chem.* 29, 1206–1215.
- (39) Syn, C. K. C. and Swarup, S. (2000) A Scalable Protocol for the Isolation of Large-Sized Genomic DNA within an Hour from Several Bacteria. *Anal. Biochem.* 278, 86–90.

---

---