CAPTURING NATURAL PRODUCT BIOSYNTHETIC PATHWAYS FROM UNCULTIVATED SYMBIOTIC BACTERIA OF MARINE SPONGES THROUGH METAGENOME MINING: A MINI-REVIEW

Agustinus Robert Uria

Research and Development Center for Marine and Fisheries Product Processing and Biotechnology,
Jalan KS. Tubun Petamburan VI, Slipi, Central Jakarta 10260, Indonesia.

Abstract

Symbiotic bacteria associated with marine sponges have frequently been proposed as the true producer of many bioactive natural products with potent anticancer activities. However, the majority of these complex symbiotic bacteria cannot be cultivated under laboratory conditions, hampering efforts to access and develop their potent compounds for therapeutic applications. Metagenome mining is a powerful cultivation-independent tool that can be used to search for new natural product biosynthetic pathways from highly complex bacterial consortia. Some notable examples of natural products, in which their biosynthetic pathways have been cloned by metagenome mining are onnamide A, psymberin, polytheonamides, calyculin, and misakinolide A. Subsequent expression of the pathways in easily culturable bacteria, such as *Escherichia coli*, could lead to the sustainable production of rare promising natural products. This review discusses principles of metagenome mining developed to gain access to natural product biosynthetic pathways from uncultured symbiotic bacteria of marine sponges. This includes detecting biosynthetic genes in sponge metagenome, creating large metagenomic library, rapid screening of metagenomic library, and clone sequencing. For many natural products made by modular polyketide synthases (PKSs) and hybrids with non-ribosomal peptide synthetases (NRPSs), their biosynthetic pathways as well as structures of final products can be predicted with high accuracy through bioinformatic analysis and sometimes combined with functional proof. Further metagenome sequencing integrated with single-cell analysis and chemical studies could provide insights into the remarkable biosynthetic capacity of uncultivated bacterial symbionts, thereby facilitating the discovery and sustainable production of a wide diversity of sponge-derived complex compounds.

Keywords: marine sponges, natural product biosynthetic pathway, uncultured symbiotic bacteria, metagenome mining

1. Introduction

Oceans are a wealthy repository of biodiversity with fourteen of 35 animal phyla being exclusively marine (Briggs, 1994; Sala & Knowlton, 2006). The highest biodiversity is found in coral reefs, where the number of species per unit area is predicted to exceed that of tropical rainforests (Haefner, 2003; Sogin et al., 2006). Coral reefs provide habitats for a wide variety of soft-bodied invertebrates, such as sponges, bryozoans, and tunicates. These marine invertebrates contain structurally diverse natural products (Hay & Fenical; 1996, Hunt & Vincent, 2006) that play an ecological role to chemically defend themselves against environmental factors, such as predation, overgrowth and settlement of other organisms, or competition for space (Proksch, 1994; Haefner, 2003). Compounds that act as chemical defense have recently become the main focus of great interest in drug development (Uria, 2012).

So far more than 20,000 compounds have been isolated from various marine organisms (Blunt et al., 2012). Among marine invertebrates, sponges are the leading sources of potent natural products in terms of the number of biologically active compounds discovered so far (Leal et al., 2012; Blunt et al., 2014). Between 1958 and 2008, around 3500 novel natural products were identified from marine sponges, including anticancer compounds, anti-inflammatory agents, and antibiotics (Hu et al., 2011). Some sponge-derived compounds have successfully been developed as clinically used drugs. This is best exemplified by two bioactive nucleosides, spongothymidine and spongouridine from the sponge *Crytotethia crypta*

*Corresponding author.
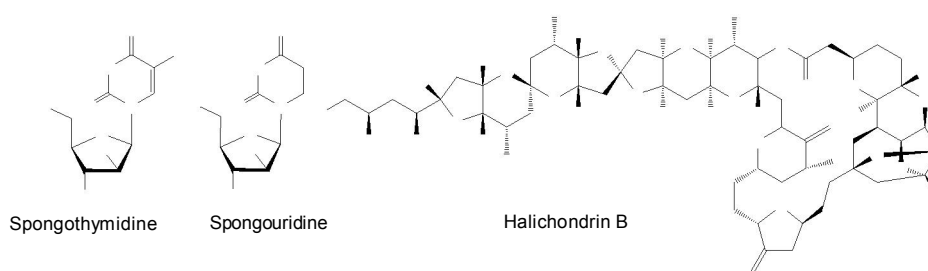E-mail: agustinus.uria@gmail.com

Figure 1. The sponge-derived compounds that inspired the synthesis of clinically used medicines: spongothymidine and spongouridine from the sponge *Crytotethia crypta* (Bergmann & Feeney, 1950, 1951; Bergmann & Swift, 1951), and halichondrin B from the Japanese sponge *Halicondria okadai* (Uemura et al., 1985).

(Bergmann & Feeney, 1950, 1951; Bergmann & Swift, 1951), which inspired the synthesis of the antiviral agent vidarabine (Ara-A), the anticancer drug cytarabine (Ara-C) (Proksch et al., 2002), and the antiviral agent zidovudine (AZT) (Bowling et al., 2007). In more recent year, the United State Food and Drug Administration (FDA) and European Medicines Agency (EMA) have approved Eribulin Mesylate (E7389, Halaven®) as a new chemotherapy agent for metastatic breast cancer patients (Donoghue et al., 2012; Gourmelon et al., 2012; Ortega & Cortés, 2012). This anticancer drug is a synthetic analog of halichondrin B from the Japanese sponge, *Halicondria okadai* (Uemura et al., 1985; Hirata & Uemura, 1986) (Figure 1).

Although a large number of novel bioactive compounds have been discovered from marine sponges, their development process into marketable drugs have been very slow, mainly due to the extremely low quantities of the substances in the sponge tissues (Uria & Piel, 2009). Compound supply by collecting a large number of a rare sponge species from its natural habitat is not a good option due to ecological and biodiversity concerns. Chemical synthesis in another hand can be very expensive, especially when structurally complex compounds are concerned (Thakur & Müller, 2004; Mohamed et al., 2008). The low yields of sponge-derived compounds, known as supply problems, are usually the major bottleneck in marine drug development (Munro et al., 1999; Sagar et al., 2010). Therefore, powerful approaches to overcome this supply problem need to be developed (Dewi et al., 2008).

Recent advances in molecular biology and bioinformatics have led to the birth of metagenomics, a powerful approach for gaining access to individual enzyme-encoding genes from uncultured marine microorganisms (Uria et al., 2005). Further development of this approach with recent advances in Deoxyribose Nucleic Acids (DNA) sequencing technologies have enabled the discovery of new biosynthetic pathways of pharmaceutically important natural products from uncultured sponge bacteria (Piel et al., 2004a; 2004b). Subsequent expression of the pathways in easily culturable bacteria, such as *E. coli*, could lead to the sustainable production of rare promising natural products. Rational protein design might permit pathway engineering, thereby creating novel pharmaceutical compounds with largely predictable structures (Uria & Piel, 2009). The standard metagenome mining procedure developed to capture biosynthetic pathways consists of the following steps: (i) obtaining sponge metagenome and molecular diagnostic of key biosynthetic genes, (ii) cloning of sponge metagenome, (iii) bioinformatics analysis (Hrvatin & Piel, 2007; Uria & Piel, 2009; Gurgui & Piel, 2010; Uria, 2012), and metagenome sequencing in combination with single-cell analysis of an enriched bacterial population from a sponge (Wilson et al., 2014). This review discusses the steps above and solutions to the challenges faced in each step are discussed in this review.

2. Bacterial Origin of Sponge-derived Natural Products

The sponge body consists of mainly three layers: (i) pinacoderm (an outer layer), (ii) choanoderm (a monolayer of flagellated collar cells that line the hollow interior atrium), and (iii) mesohyl (the connective tissue layer located between the pinacoderm and the choanoderm (Ruppert et al., 2004; Taylor et al., 2007). In general, the pinacoderm is typically dominated by photosynthetic bacteria such as cyanobacteria because it is exposed to sunlight. The mesohyl is primarily populated by heterotrophic and autotrophic bacteria (Kennedy et al., 2007) (Figure 2). For some well-studied marine sponges, symbiotic microbial consortia can be very complex with at least 18 bacterial and archaeal phyla (Taylor et al., 2007). The symbiotic microbial density may account up to 60% of the total sponge biomass with 6.4×10^8 to 1.5×10^9 cells per ml of sponge extract (Scheuermayer et
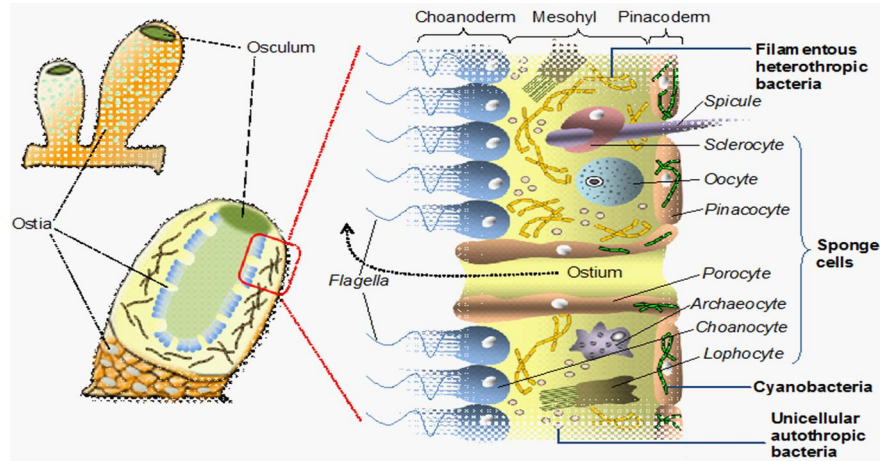
Figure 2. The sponge's tissue containing symbiotic bacterial cell types. Redrawn from Uria (2012) that was modified from Ruppert et al. (2004).

al., 2006) or up to 10^8 -10^{10} bacterial cells per gram of body wet weight (Hentschel et al., 2006). The genetic diversity or variation within and between species (Achtman & Wagner, 2008) makes the symbiotic community even more complex (Uria, 2012).

For a long time, highly complex symbiotic bacterial consortia have been proposed as the actual producers of many sponge-derived natural products (Piel, 2004). This symbiont hypothesis is supported by the frequently observed phenomena that many invertebrate-derived compounds are structurally similar to typical bacterial compounds (Bewley & Faulkner, 1998; Piel, 2004), and marine invertebrates often contain complex compound families commonly known from bacteria (Piel, 2006; Uria & Piel, 2009). If the symbiotic bacteria could be cultivated under laboratory conditions, then fermentation may facilitate the sustainable production of drug candidate at the industrial scale (Uria & Piel, 2009). However, the majority of these symbiotic bacteria are not amenable to cultivation outside their hosts (Friedrich et al., 2001; Webster & Hill, 2001), hampering efforts to study and access potential marine natural products. Study of the microbial diversity in the Great Barrier Reef sponge *Rhopaloeides odorabile*, for example, revealed that the culturable percentage of symbiotic bacteria accounts for only 0.1-0.23% of the total bacterial community (Webster & Hill, 2001).

Early studies of natural products in uncultured symbiotic bacteria involved mechanical separation of bacterial cell types, followed by chemical analysis of each cell type (Unson & Faulkner, 1993; Unson et al., 1994), and subsequent 16S-ribosomal Ribose Nucleic Acids (rRNA) gene-based phylogenetic analysis of cell types to determine the taxonomic status (Schmidt et al., 2000; Uria & Piel, 2009). A number of cell separation methods have been used to obtain bacterial

cells from hosts, including a fluorescence activated cell sorter (FACS) (Unson et al., 1994), differential centrifugation (Bewley et al., 1996; Piel et al., 2004; Wilson et al., 2014), percoll density gradient (Ridley, et al., 2005), simple squeezing (Schmidt et al., 2005), and micromanipulation (Grindberg et al., 2011). Bewley et al. (1996) used differential centrifugation method to separate three bacterial cell types from the Palauan sponge *Theonella swinhoei* (Figure 3). The isolated bacterial cell types are (i) unicellular cyanobacterial fraction identified as *Aphanocapsa feldmanni*, (ii) filamentous heterotrophic bacteria, and (iii) unicellular heterotropic bacteria (Bewley et al., 1996). By 16S-rRNA gene analysis, Schmidt et al. (2000) showed the taxonomic status of such uncultured filamentous bacteria as Candidatus "*Entotheonella palauensis*".

Chemical analysis of the cell fractions using [1]H-Nuclear Magnetic Resonance (NMR) spectroscopy in combination with Transmission Electron Microscopy (TEM) showed the accumulation of swinholide A in the mixed unicellular bacterial fraction. Whereas the antifungal theopalaumide was present mainly in the filamentous bacteria (Bewley et al., 1996). In a separate study, Brück et al. (2008) employed Fluorescent *In Situ* Hybridization (FISH) to show the presence of the symbiotic "*Entotheonella* sp." obtained from the Caribbean Sea sponge, *Discodermia dissolute* (Brück et al., 2008). However, such chemical localization alone cannot give convincing proof for the bacterial origin due to the possibility that natural products move across cellular membranes and whole tissues (Uria & Piel, 2009).

Investigating the biosynthetic origin of sponge-derived natural products at the genetic level is extremely challenging due to the complexity of sponge metagenome that contains various homologous genes from diverse pathways (Piel, 2010). The fact that
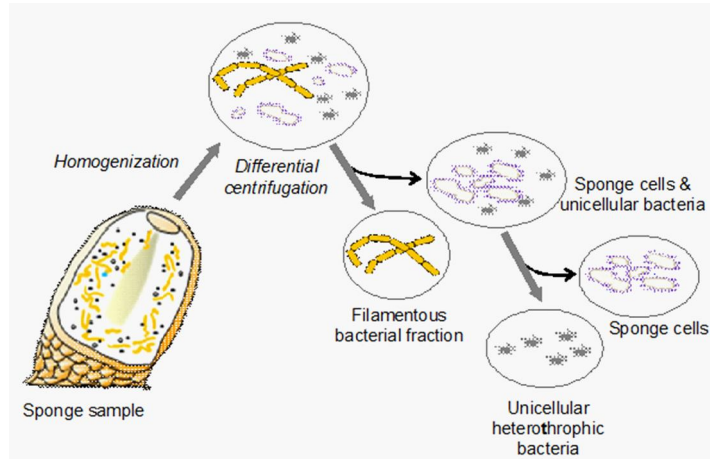
Figure 3. Mechanical separation of bacterial cells from sponge tissue. The sponge sample is initially homogenized in calcium/magnesium-free artificial seawater (CMF-ASW), and then subjected to differential centrifugation to separate several cell types based on different size and densities (Uria, 2012). Redrawn from Uria (2012).

biosynthetic pathways leading to pharmaceutically important natural products are typically clustered in bacterial chromosomes (Kelinkauf & Döhren, 1990; Kim et al., 2010) has greatly facilitated the isolation of biosynthetic gene clusters from complex microbial consortia without prior cultivation (Uria & Piel, 2009). Using metagenomic approaches, Piel and his colleagues cloned a gene cluster for the biosynthesis of onnamine/theopederin from the Japanese sponge *T. swinhoei* (Piel et al., 2004b). The putative onnamide gene cluster showed a typical bacterial architecture, such as tightly packed genes, lack of introns and polyadenylation sites, and be preceded by putative Shine-Dalgarno (Piel et al., 2004b). This provides the first genetic evidence for the bacterial origin of marine invertebrate-derived natural products (Figure 4).

Some other examples of sponge-derived natural products, in which their biosynthetic pathways have been characterized include psymberin from the sponge *Psammocinia* aff. *bulbosa* (Fisch et al., 2009), polytheonamides A and B from Japanese *T. swinhoei* chemotype yellow (Freeman et al., 2012) (Figure 4), calyculin from *Discodermia calyx* (Wakimoto et al., 2014), and misakinolide A (Uria, 2012; Ueoka et al., 2015). Polytheonamides A and B are giant peptides that function as unimolecular pores in cell membranes (Iwamoto et al., 2010). The misakinolide biosynthetic pathway harbors an additional module that match the structures of swinholide A (Carmely & Kashman, 1985; Kitagawa et al.,1990; Doi et al., 1991) and hurghadolide A (Youssef & Mooberry, 2006). Discovery of such biosynthetic pathways provides not only convincing proof for the bacterial origin of sponge-derived natural products but also facilitates the sustainable supply of rare promising complex compounds.

Although the bacterial origin of invertebrate-natural products can be verified at the genetic level through metagenome mining, this approach alone cannot provide specific taxonomic information of the natural product producers. By integrating metagenome sequencing and single-cell analysis, Piel and his colleagues pinpointed a single bacterial phylotype of the *candidatus* genus 'Entotheonella' that makes almost all toxic compounds (more than 40) isolated from the Japanese sponge *Theonella swinhoei* Y (Wilson et al., 2014). Interestingly, this uncultured bacterial phylotype is widely distributed in taxonomically diverse sponge species collected at 20 different locations and belongs to a new environmental taxon proposed as candidate phylum 'Tectomicrobia' (Wilson et al., 2014). Wakimoto et al. (2014) has recently reported the biosynthetic pathway of calyculin from *D. calyx*. This cytotoxic compound represents a form of activated chemical defense in response to sponge host injury. Catalysed Reported Deposition (CARD)-FISH probing supported that calyculin biosynthesis occurs in "Entotheonella sp." (Wakimoto et al., 2014).

3. Detection of Biosynthesis Genes in Sponge Metagenome

The presence of key biosynthesis genes in marine sponges can be detected by Polymerase Reaction Chain (PCR) cloning using total DNA from a sponge sample. PCR cloning procedure usually involves the following steps: designing PCR primers to target a small conserved region in a biosynthesis gene, PCR-amplification of the conserved region, cloning of PCR products into *Escherichia coli*, DNA sequencing, and
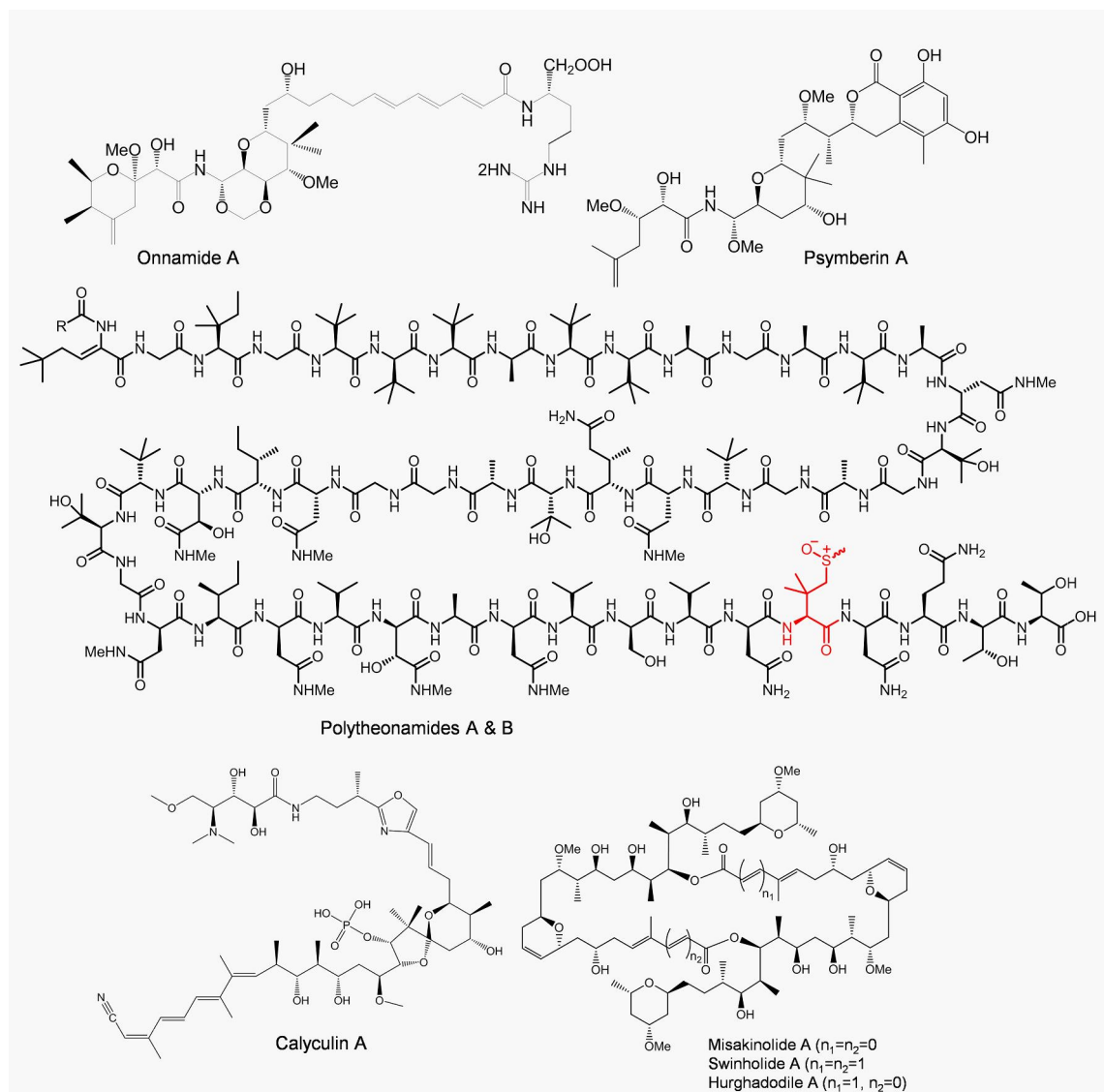
Figure 4. Sponge-derived natural products, in which their biosynthetic pathways have been cloned by metagenome mining. These include onnamide A (Piel et al., 2004b), psymberin (Fisch et al., 2009), and polytheonamides A & B that differ in the configuration of the sulfoxide moiety indicated by red color (Freeman et al., 2012), calyculin A (Wakimoto et al., 2014), and misakinolide A (Uria, 2012; Ueoka et al., 2015).

bioinformatic analysis. Results obtained by this PCR cloning could provide preliminary insights into diversity of potential genes involved in the biosynthesis of sponge-derived natural products.

Since many potent antitumor compounds isolated from marine sponges belong to complex polyketides (PKs) and hybrids with non-ribosomal peptides (NRPs) (Piel, 2004; Fusetani, 2010), PCR cloning is mostly directed to target genes involved in the biosynthesis of these two compound classes. This section briefly describes the biosynthetic machinery of PK and NRP, and then explain which parts of PK/NRP biosynthetic pathways can be used as targets by PCR cloning.

Biosynthesis of PK is catalyzed by multifunctional proteins termed polyketide synthases (PKS) (Llewellyn & Spencer, 2007). Mechanisms of how PKS systems work have comprehensively been reviewed by Hertweck (2009) and Piel (2010). In principle, PKS systems consist of modules. Each module contain at least these three main domains: a ketoacyl synthase (KS) domain catalyzes the chain elongation, an acyltransferase (AT) domain chooses the correct acyl-CoA extender, and an acyl carrier protein (ACP) domain acts as anchor to the polyketide chain (Cane et al., 1998). Other optional domains may be present to perform various functional modifications of the polyketide backbones, producing β-hydroxyl, olefinic
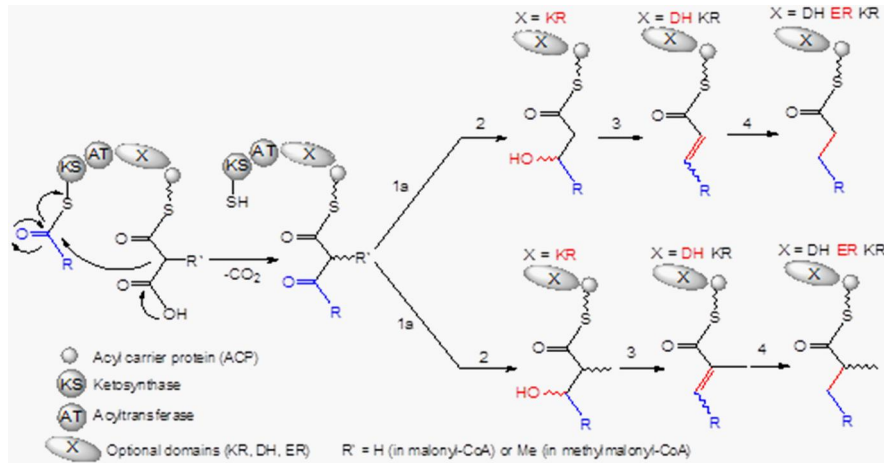
Figure 5. Basic biosynthetic steps in modular PKS, including extension with malonyl-CoA (1a) or extension with methylmalonyl-CoA (1b), keto-reduction (2), dehydration (3), and double-bond reduction (4). Abbreviations: KR, *β*-ketoacyl reductase; DH, dehydratase; and ER, enoyl reductase. Redrawn from Uria (2012).
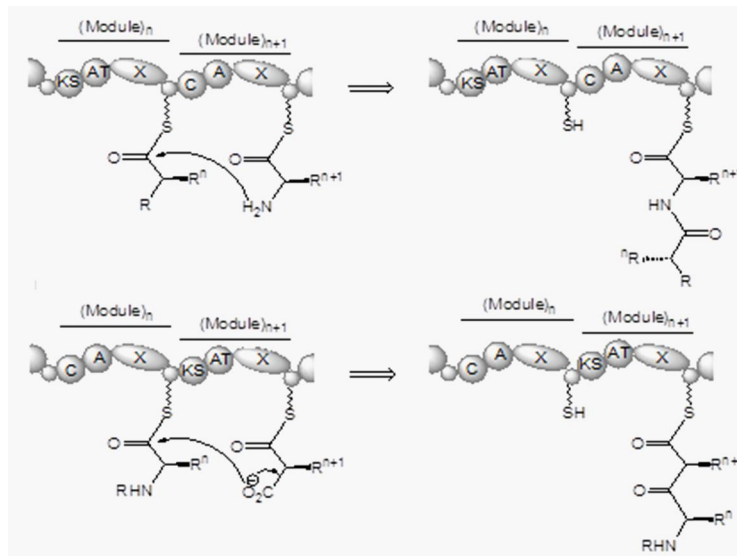

Figure 6. Bond formation in NRP and hybrids with PK: C-N formation in the PKS/NRPS hybrid (on the top), and C-C formation in the NRPS/PKS hybrid (below). Abbreviations: AT, acyl transferase; ACP, acyl carrier protein; KS, β-ketoacyl synthase; C, condensation domain; A, adenylation domain; 0 or T, thiolation domain; x, optional domain(s). Redrawn from Uria (2012).

and fully saturated moieties (Cane et al., 1998; Rawlings, 2001; Tsai & Ames, 2009) (Figure 5).

In peptide biosynthesis, NRPSs use a similar strategy as PKSs. Each module in NRPS system incorporate and modify an amino acid unit into the growing peptide chain (Du et al., 2001). There are at least three domains in a typical NRPS module, namely (i) an adenylation (A) domain that serves to activate and incorporate amino acid, (ii) a thiolation (T) domain that performs thioesterification of the activated residue, and (iii) a condensation (C) domain that extends of the growing peptide chain via

transpeptidation (Schwarzer & Marahiel, 2001; Schwarzer et al., 2003). In the biosynthesis of PK/NRP hybrids, C domain mediates C-N bond formation through nucleophilic substitution between the acyl group of the growing polyketidyl chain from the upstream PKS module and the amino group of the downstream aminoacyl-*S*-T (Du et al., 2001) (Figure 6). At the NRPS/PKS interface, C domain catalyzes C-C bond formation by transferring the growing intermediate peptidyl-*S*-T at the upstream NRPS module to the residue of the KS active site (cystein), thereby generating a peptidyl-*S*-KS intermediate. The

decarboxylative condensation of the resulting intermediate peptidyl-*S*-KS with a malonyl-*S*-ACP on the downstream PKS module elongates the peptide chain with a short carboxylic acid (Fischbach & Walsh, 2006; Giessen & Marahiel, 2012). Among domains present in NRPS modules, an adenylation (A) can be chosen as the cloning target due to the highly conserved amino acid sequences throughout this domain in comparison with other NRPS domains (Turgay et al., 1992; Turgay & Marahiel, 1994).

Among domains present in PKS systems, KS domains are usually used as the diagnostic target, because sequences of this domain are most conserved sequences compared with other PKS domains (Aparicio et al., 1996). PCR-amplification of the KS regions from PKS gene clusters in total sponge DNA is very challenging due to the huge diversity of homologous genes from various pathways present in a complex sponge symbiosis (Piel, 2010). The huge amounts (86% in average) of fatty acids in the sponges makes it difficult to detect biosynthetic genes for complex polyketide (Fieseler et al., 2007). PCR detection of PKS genes in the total sponge DNA using degenerate primers showed that none of the obtained KS amplicons involved in the biosynthesis of complex bioactive polyketides (Hochmuth et al., 2010; Wilson et al., 2014). Instead, they belonged to KS domains of SupA (Uria, 2012; Wilson et al., 2014) that harbors a single complete PKS module (KS-AT-DH-MT-ER-KR-ACP) with additional KS and inactive AT domains at the C-terminus. This domain architecture is typical animal FASs (Schweizer & Weissman, 2001), in which the product backbone is fully reduced, thereby resulting in a methyl-branched fatty acid (Fieseler et al., 2007).

To solve the problem, the complex microbial consortia was simplified using differential centrifugation method to obtain the filamentous bacterial fraction mainly dominated by large filamentous cells *Candidatus* "Entotheonella sp." (Piel et al., 2004a; Uria, 2012; Wilson et al., 2014). PCR cloning using such symbiotic cell pellet resulted in KS amplicons involved in complex polyketide biosynthesis (Wilson et al., 2014).

To detect the PKS/NRPS machinery of a known compound, in which the encoding DNA sequence data is unknown, degenerate primers to target such machinery could be designed based on the predicted biogenesis of the compound or its building blocks. For example, degenerate primers to target the biosynthetic pathway of theonellamide F, an antifungal cyclic peptide from *T. swinhoei* (Matsunaga et al., 1989), could be based on the proposed biogenesis of such peptide and its unusual amino acids, as described by Uria (2012) in Figure 7. Theonellamide F contains 10 amino acids, in which three of them are unusual amino acids, namely a brominated phenylalanine (BrPhe), (3*S*,4*S*,5*E*,7*E*)-3-amino-4-hydroxy-6-methyl-8-(*p*-bromophenyl)-5,7-octadienoic acid (Aboa), and (2*S*,3*R*)-3-hydroxyasparagine-(2*S*,4*R*)-2-amino-4-hydroxyadipic acid (α-AhAA) (Matsunaga et al., 1989). Incorporation of all of the amino acids into theonellamide F is predicted to be enzymatically catalyzed by NRPS modules and a PKS module. Formation of Aboa might involve a PKS module that elongates and modifies the L-Phe or BrPhe installed on a NRPS module. This might occur through a similar mechanism reported for 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid (Adda) in microcystins biosynthesis (Namikoshi
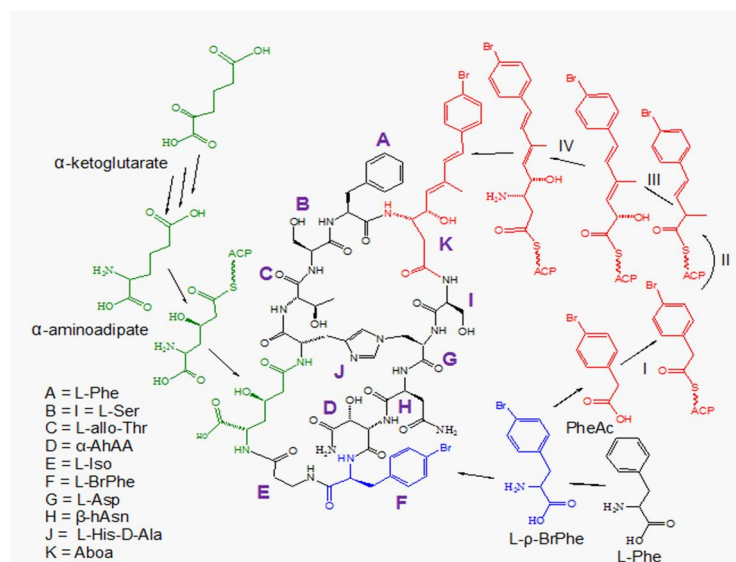

Figure 7. The proposed biogenesis of three unusual residues in theonellamide F: BrPhe (blue color), *α*-AhAA (green color), and Aboa (red color). Redrawn from Uria (2012).

et al., 1992; Tillet et al., 2000). As described in Figure 7, the installed L-Phe or BrPhe would undergo KS-catalyzed elongation (I) and a series of modifications, including a double bond shift by β,γ-dehydration and α-methylation (II), α-hydroxylation (III), and aminotransferase (IV). This suggests that degenerate primers targeting the theonellamide biosynthetic pathway could be designed on the basis of KS domain of PKS and A domain of NRPS (Uria, 2012).

To specifically target the biosynthetic pathways of theonellamides, degenerate primers could be based on the biogenesis of BrPhe and α-AhAA. As shown in Figure 7 (Uria, 2012), the proposed biogenesis of the incorporated BrPhe is presumably from L-Phe via enzymatic halogenation, as has been shown for tryptophan-5-halogenase in the pyrroindomycin B biosynthesis (Zehner et al., 2005). This suggests that designing diagnostic PCR-primers could be based on the halogenase enzyme. The α-AhAA probably arises from the hydroxylation of L-α-aminoadipic acid, which is an intermediate in the lysine biosynthetic pathway (Velasco et al., 2002). Incorporation of L-α-aminoadipic acid into the structure of theonellamide might be catalyzed by a NRPS assembly line, as reported for the biosynthetic pathway of the tripeptide d-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV), a common intermediate in the penicillins and cephalosporins (Martín, 2000). This suggests that the A domain specific for aminoadipic acid could be used as a basis for primer design (Uria, 2012).

4. Cloning of Sponge Metagenome

Capturing natural product biosynthetic pathways requires a large metagenomic library that represent microbial genomes present in the sponge system. A complex metagenomic library is usually constructed by isolation of high molecular weight (HMW) DNA from sponge sample, followed by cloning of HMW DNA fragments into *E. coli* (Uria & Piel, 2009). A huge amount of polysaccharides present in the sponge tissue (Bucior & Burger, 2004; Esteves et al., 2011) may interfere the isolated DNA and inhibit enzymatic reactions in DNA modifications (Kaufman et al., 1999). Therefore, polysaccharide removal is necessary in metagenomic library construction. Cetyltri methylammonium bromide (CTAB) has been proven to be useful in removing polysaccharides from the sponge cell extract (Piel et al., 2004; Schirmer et al., 2005). In combination with a high concentration of urea has allowed cell wall lysis (Hrvatin & Piel 2007; Gurgui & Piel, 2010).

Another way to remove the excessive polysaccharides is homogenization of a sponge sample, followed by differential centrifugation to collect symbiotic bacterial cells. Subsequent treatment of the obtained cell pellet with lysozyme and achromopeptidase allowed cell lysis, thereby releasing the HMW bacterial DNA (Bertrand et al., 2005; Uria et al., 2007). HMW DNA fragments of interest (*e.g.* approximately 35-40 kilo base pairs/kb) can be selected and purified by gel electrophoresis, followed by either enzymatic gel extraction using GELase (Epicentre) or column-based gel extraction in order to remove the remaining contaminants including RNA. GELase contains an enzyme capable of digesting α-agarose, allowing the Low Melting Point (LMP) gel's carbohydrate backbone to break down into soluble small oligosaccharides (Epicentre; Uria, 2012).

For cloning of large-sized inserts (35-40 kb), a high-capacity vector with low copy number is often used in order to prevent the instability and rearrangement of the cloned large fragments (Kim et al., 1992; Shizuya et al., 1992). An example is the fosmid pCC1FOS (Epicentre), in which its low copy number is controlled by the F factor system that keeps 1-2 copies per cell. The pCC1FOS copy number can be increased to 10-50 copies per cell in the growing media supplemented with L-arabinose (Epicentre, Wild et al., 2002; Westernberg et al., 2010). The detailed procedure of how to construct a fosmid-based metagenomic library is described by Gurgui & Piel (2010) (Figure 8).

When using a dephosphorylated linear fosmid, one of important steps in the library construction is repairing the DNA fragment termini prior to the blunt legation with a fosmid vector, because during extraction and purification, the DNA fragment ends could be fray or not blunt without phosphate residues. T4 DNA polymerase is usually used to convert the frayed ends into blunt ends by filling the recessed 3' termini and removing the protruding 3'-termini. Subsequently, terminal phosphotransferase is used to convert the hydroxyl group at the 3'-termini into a phosphate group (Brown, 2006). Metagenomic library can be generated in 3-D format in semi-liquid medium, as recently reported by Hrvatin & Piel (2007). This library format allows growing well-separated clones at high densities between 1000 to 4000 clones/ml (Uria, 2012).

All the clones in 3-D format library are usually organized into pools. Each pool contains approximately 1000-4000 clones/ml semi-liquid medium (Hvartin & Piel, 2007; Uria, 2012). Individual pools can be added with glycerol, mixed well by vortex and stored at -80 °C for a long time. This library format facilitates rapid screening of a complex metagenomic library by whole-cell PCR and pool dilution (Hvartin & Piel, 2007). For library screening, a small aliquot of all individual pools are initially combined into super-pools. Individual super-pools are screened by whole-cell PCR. Each pools located in the same positive
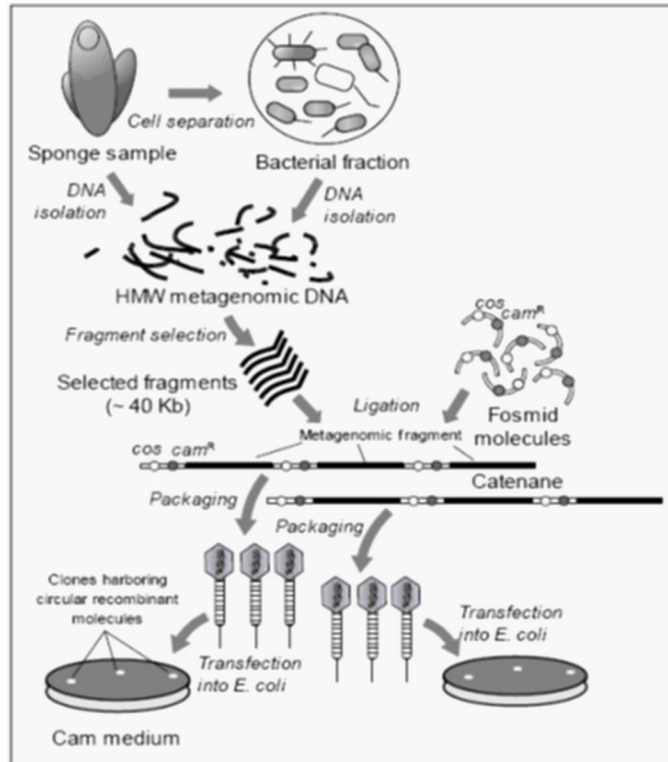
Figure 8. Metagenome mining protocol may include isolation of HMW DNA from either a sponge sample or its symbiotic bacterial cells. After being separated by electrophoresis on a LMP agarose gel, the DNA fragments of ~35-40 kb are extracted from the gel, repaired at both termini, ligated with dephosphorylated fosmids, packed in bacteriophage particles that are then used to transfect *E. coli.* The resulting library is then grown in 3-D format in semi-solid media and screened by whole cell PCR (Uria, 2012). Redrawn from Uria (2012).

super-pool are checked by PCR. Positive clone pools are then diluted at lower cell density and re-screened by PCR. This combination of cell dilution and PCR screening is repeated. When the clone density is around 20 cfus/ml, the cell dilution were plated and grown at the colony density of around 200 cfus/ plate. Colony PCR can then be carried out to isolate a single positive clone (Hvartin & Piel, 2007; Gurgui & Piel, 2010; Uria, 2012).

If a biosynthetic pathway is too large, it requires isolation of more than one overlapping fosmid clones isolated from the library. Finding the remaining parts of the partial gene cluster can be performed by the homology-based screening using specific primers that specifically recognize the target biosynthetic pathway. Specific primers are usually designed based on the end sequences of the previously isolated fosmid and employed to screen the additional portions of the gene cluster. This process is repeated until the entire gene cluster is isolated. Clone analysis can be initiated with end-sequencing and sometimes followed by sub-cloning. If sub-cloning indicates that the isolated clones are promising, then complete sequencing of individual

clones should be pursued. Bioinformatic analysis of the sequenced clones could provide insights into the biosynthetic pathways and the encoded compounds (Uria, 2012).

5. Bioinformatic Analysis and Functional Proof

A positive clone isolated from a metagenomic library needs to be analyzed further in order to know whether it harbors a biosynthetic pathway of interest or not. Obtaining the functional proof by heterologous expression of the entire biosynthesis gene cluster is sometimes difficult because the cluster is very large located in more than one clones. In such case, bioinformatics analyses of the cloned DNA sequences are very useful to predict the encoded biosynthetic pathways and even the chemical structure of final products. This involves the use of various softwares or online programs to analyze the sequence data. For example, Seqman NGen® (DNASTAR, Madison, WI) and Geneious (Drummond et al., 2011) can be used to assemble the sequencing reads. The online-accessible software FramePlot (Ishikawa & Hotta, 2006) can be used to identify potential open reading

frames (ORFs). Sometimes the gene cluster function predicted using bioinformatics needs to be experimentally proved by heterologous expression and activity assay of just one or some small enzymes/ proteins as a part of the entire giant biosynthetic pathway. In such case, substrates required for enzymatic/protein analyses should be provided either by purchasing commercially available substrates or through chemical synthesis of predicted substrates.

For modular PKS system or its hybrid with NRPS system, the module architecture including the domains present in each module can be predicted using some online programs, such as Basic Local Alignment Search Tool (BLAST) analysis (Altschul et al., 1990; McGinnis & Madden, 2004), and PKS/NRPS analysis Web-site (Bachmann & Ravel, 2009). Other tools that can be used to analyze large DNA sequence data for identifying PK and NRP biosynthetic pathways include Antibiotics and Secondary Metabolite SHell (antiSMASH) (Medema et al., 2011; Blin et al., 2013), Natural Product Domain Seeker (NaPDoS) (Ziemert et al., 2012), and NP.searcher (Li et al., 2009). Subsequently based on the module architecture and substrate specificity, the PK/NRP intermediates and biosynthetic pathways could be predicted accurately (Hertweck, 2009; Piel 2010).

As described in section 3, the A domains in the NRPS pathways function to select/activate and incorporate amino acid residues (Schwarzer et al., 2003). The key residues responsible for coordination of Adenosine Triphosphate (ATP) and substrate, also known as nonribosomal codes, for the substrate specificity of the A domains have been established (Challis et al., 2000). The nonribosomal codes of A domains could be revealed using the online-accessible programs, such as NRPSpredictor2 (Röttig et al., 2011) and NRPSsp (Prieto et al., 2012). In certain cases, functional proof for the substrate specificity of A domains needs to be provided by their overexpression in *E. coli*, protein purification and subsequent activity analysis towards the predicted substrates, for example, using mass exchange-based adenylation assay (Phelan et al., 2009; Höfer et al.,2011; Crüsemann et al., 2013; Wilson et al., 2014).

In PK pathways, the selection of the correct building blocks is performed by AT domains. Structural studies of AT's binding pocket have been used to create predictive tools for the specificity of malonyl-CoA (MCoA), methylmalonyl-CoA (MMCoA), and other less common substrates (Anand et al., 2010; Yadav et al., 2003; Boddy, 2014). By predicting the AT's substrate specificity and the order of other PKS domains, the PK biosynthetic pathway could be defined that follow the collinearity rule (Hertweck, 2009; Boddy 2014).

However, for PKS usually characterized by the lack of AT domains in each module, called *trans*-AT PKS, predicting the biosynthetic pathways and the final product is more challenging, because it cannot be predicted straightforward just based on the domain organization that follow the collinearity rule (Piel, 2010). To solve the problem, the Piel group has recently developed a phylogeny-based approach to predict the biosynthetic pathway and the final product structure of an AT-lacking PKS system with high confidence based on KS sequence and module architecture (Nguyen et al., 2008, Piel, 2010; Teta et al., 2010; Uria, 2012). For *trans*-AT PKSs containing NRPS modules, thereby resulting NRPS-PKS hybrids, the specificity of the KS domains towards NRPS-derived amino acid intermediates could functionally be tested using a Mars Spectometry-based method (Jenner et al., 2013; Kohlhaas et al., 2013).

6. Metagenome Sequencing

Considering the extremely complexity of sponge microbes (Hentschel et al., 2002; Scheurmayer et al., 2006; Taylor et al., 2007) and the fact that the majority of them are not amenable to cultivation (Amann et al., 1995; Hugenholtz et al., 1998), sequencing of entire microbial communities present in a sponge specimen is very challenging to accomplish (Uria & Piel, 2009). To tackle the challenge, Piel and his colleagues initially enriched a population of large filamentous bacteria by differential centrifugation. Subsequent application of fluorescence-assisted cell sorting (FACS), multiple displacement amplification (MDA), and diagnostic PCR techniques enabled Piel and his colleagues to pinpoint "Entotheonella" as the producer of onnamide-type compounds and polytheonamides (Wilson et al., 2014). The biosynthetic pathways of these two compound types have previously been described by Piel (et al., 2004b) and Freeman et al. (2012), respectively.

Subsequently Piel and his colleagues conducted metagenome sequencing of the enriched filamentous symbiotic bacteria by integrated Illumina, 454-pyro, PacBio and Sanger sequencing methods. Bioinformatic analysis of the sequence data revealed the biosynthetic capacity of two dominant "Entotheonella" variants designed as TY1 and TY2. In particular, the chromosome TY1 of more than 9 Mb in size was found to contain a huge number of PK and NRP gene clusters, in which some of them encode the biosynthesis of known bioactive peptides, such as cyclotheonamides, keramamides and nazumamide A. This finding was supported by high-resolution MS analysis of the sponge and filamentous cell extracts as well as functional studies of two

selected NRPS adenylation domains. Whereas the TY2 chromosome contained gene clusters for two PKs, at least two NRPs, and a proteusin (Wilson et al., 2014). Interestingly, onnamide and polytheonamide gene clusters were located in plasmid-derived contigs (Wilson et al., 2014). Therefore, metagenome mining in combination with single-cell, functional and chemical analyses conducted by Piel and his colleagues revealed the remarkable biosynthetic potential of the uncultivated bacterial symbiont `Entotheonella factor TY1` (Wilson et al., 2014), thereby offering exciting opportunities for the discovery and sustainable supply of marine drugs.

7. Conclusion

Metagenome mining involves molecular cloning and bioinformatic analysis of genes encoding the biosynthesis of drug candidates from uncultured microbial consortia. Application of this approach to find a target biosynthetic pathway from a highly complex sponge symbiosis require (i) detecting biosynthesis genes in sponge metagenome, (ii) cloning of sponge metagenome and clone sequencing, and (iv) bioinformatic analysis and functional proof. Development of this approach enables cloning of pharmaceutically important biosynthetic pathways from complex microbial consortia in marine sponges without prior cultivation. Further metagenome sequencing integrated with single-cell, chemical and functional studies could provide insights into the remarkable biosynthetic capacity of uncultivated bacterial symbionts of marine sponges. Discovery of these biosynthetic pathways provides a basis for the sustainable production of promising sponge-derived complex compounds.

References

Achtman, M., & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat. Rev.,* 6;431-440.

Altschul, S.F., Gish, W., Miller, W. Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.,* 215;403-410.

Amann, R.I., Ludwig, W. & Schleifer, K.-H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.,* 59(1);143-169.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z. and Mohanty, D. (2010). SBSPKS: structure based sequence analysis of polyketide synthases. *Nucl. Acids Res.* 38;W487–W496.

Aparicio, J.F., Molnar, I., Schwecke, T., König, A., Haydock, S.F., Khaw, L.E., Staunton, J. & Leadley, P.F. (1996). Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene*, 169;9-16.

Bachmann, B.O, & Ravel, J. (2009). *In Sillico* prediction of microbial secondary metabolic pathways from DNA sequence data. *Meth. Enzymol.* 458;181-217.

Bergmann, W. & Feeney, R.J. (1950). The isolation of a new thymine pentoside from sponges. *J. Am. Chem. Soc.,* 72;2809–2810

Bergmann, W. & Feeney, R.J. (1951). Contributions to the study of marine products, 32: the nucleosides of sponges. I. *J. Org. Chem.*, 16;981–987

Bergmann, W., & Swift, A.N., (1951). Contributions to the study of marine products, 30: Component acids of lipids of sponges. I. *J. Org. Chem.* 16;1206–1221

Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M. & Simonet, P. (2005). High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J. Microbiol. Meth.* 62;1-11.

Bewley, C.A. & Faulkner, D.J. (1998). Lithistid sponges: star performers or hosts to the stars. *Angew. Chem. Int. Ed.*, 37(16),2162–2178.

Bewley, C.A., Holland, N.D. & Faulkner, D.J. (1996). Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. *Experientia,* 52;716-722.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., & Weber, T. (2013). antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucl. Acids Res.* 41;W204–W212.

Boddy, C.N. (2014). Bioinformatics tools for genome mining of polyketides and non-ribosomal peptides. *J. Ind. Microbiol. Biotechnol.,* 41;443-450.

Bowling, J.J., Kochanowska, A.J., Kasanah, N., & Hamann, M.T. (2007). Nature's bounty – drug

discovery from the sea. *Expert Opin. Drug Discov.*, 2(11);1505-1522.

Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H.G., & Prinsep, M.R. (2012). Marine natural products. *Nat. Prod. Rep.*, 29;144–222.

Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H.G., & Prinsep, M.R., (2014). Marine natural products. *Nat. Prod. Rep.*, 31;160–258.

Briggs, J.C. (1994). Species diversity: land and sea compared. *Syst. Biol.*, 43(1);130-135.

Brown, T.A. (2006). Gene cloning & DNA analysis, an introduction - 5th edition. *Blackwell Publishing Ltd.*

Brück, W.M., Sennett, S.H., Pomponi, S.A., Willenz, P., & McCarthy, P.J. (2008). Identification of the bacterial symbiont *Entotheonella* sp. in the mesophyl of the marine sponge *Discodermalide. ISSM J.*, 1-5.

Bucior, I., & M.M. Burger (2004). Carbohydrate-carbohydrate interaction as a major force initiating cell-cell recognition. *Glycoconjugate J.*, 21(3-4);112-123.

Cane, D.E., C.T. Walsh, & Khosla, C. (1998). Mutations harnessing the biosynthetic code: combinations, permutations, and mutations. *Science,* 282;63-68.

Carmely, S., and Kashman, Y. (1985). Structure of swinholide A, a new macrolide from the marine sponge, *Theonella swinhoei. Tetrahedron Lett.* 26(4);511-514.

Challis, G.L., Ravel, J., & Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.*, 7;211-224.

Crüsemann, M., Kohlhaas, C., & Piel, J. (2013). Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chemical Sci.* 4,041–1045.

Dewi, A.S., Tarman, K., & Uria, A.R. (2008). Marine Natural Products: Prospects and Impacts on the Sustainable Development in Indonesia. The Indonesian Student Scientific Meeting (ISSM) on Sustainable development in Indonesia: An interdisciplinary approach, 13–15 May 2008 at Delft University of Technology.

Donoghue, M., Lemery, S.J., Yuan, W., He, K., Sridhara, R., Shord, S., Zhao, H., Marathe, A., Kotch, L., Jee, J., Wang, Y., Zhou, L., Adams, W., Jarral, V., Pilaro, A., Lostritto, R., Gootenberg, J., Keegan, P., & Pazdur, R. (2012). Eribulin Mesylate for the treatment of patients with refractory metastatic breast cancer: use of a "physicans` choice" control arm in a randomized approval trial. *Clin. Cancer Res.* DOI:10.1158/1078-0432.CCR-11-2149.

Doi, M., Ishida, T., Kobayashi, M., Kitagawa, I. (1991). Molecular conformation of swinholide A, a potent cytotoxic dimeric macrolide from the Okinawan marine sponge *Theonella swinhoei:* X-ray crystal structure of its diketone derivative. *J. Org. Chem.* 56;3629-3632.

Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., & Wilson, A. (2011). Geneious v5.5.

Du, L., Sánchez, C., & Shen, B. (2001). Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic Eng.*, 3;78-95.

Esteves, A.I.S., Nicolai, M., Humanes, M., & Goncalves, J. (2011). Sulfated polysaccharides in marine sponges: extraction methods and anti-HIV activity. *Mar. Drugs*, 9(1);139-153.

Höfer, I., Crüsemann, M., Geers, B., Flaschhaar, D., Cai, X., Zeeck, A., & Piel, J. (2011). Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. *Chem. Biol.*, 18;381–391.

Fisch, K.M., Gurgui, C., Heyke, N., van der Sar, S.A., Anderson, S.A., Webb, V.L., Taudien, S., Platzer, M., Rubio, B.K., Robinson, S.J., Crews, P., & Piel, J. (2009). Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.*, 5(7);494-501.

Fischbach, M.A. & Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.,* 106(8);3468-3496.

Fieseler, L., Hentschel, U., Grozdanov, L., Schirmer, A., Wen, G., Platzer, M., Hrvatin, S., Butzke, D., Zimmermann, K., & Piel, J. (2007). Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl. Environ. Microbiol.*, 73;2144–2155.

Friedrich, A.B., Fischer, I., Proksch, P., Hacker, J., & Hentschel, U. (2001). Temporal variation of the microbial community associated with the Mediterranean sponge *Aplysina aerophoba. FEMS Microbiol. Ecol.*, 38(2-3);105–113.

Freeman, M.F., Gurgui, C., Helf, M., Morinaka, B., Uria, A.R., Oldham, N.J., Sahl, H-G., Matsunaga, S., & Piel, J. (2012). Metagenome mining reveals ribosomal peptide biosynthesis of unprecedented structural scope. *Science*, 338;387-390.

Fusetani, N. (2010). Antifungal peptides in marine invertebrates. *ISJ* 7,53-66.

Giessen, T.W. & Marahiel, M.A. (2012). Ribosome-independent biosynthesis of biologically active peptides: application of synthetic biology to generate structural diversity. *FEBS Lett.*Doi:10.1016/j.febslet.2012.01.017

Gourmelon, C., Frenel, J.S. & Campone, M. (2012). Clinical evidence for the role of eribulin mesylate in the treatment of breast cancer. *Clin. Investig.,* 2(2),207-213.

Grindberg R.V., Ishoey, T., Brinza, D., Esquenazi, E., Coates, R.C., Liu, Wt., Gerwick, L., Dorrestein, P.C., Pevzner, P., & Lasken, R. (2011). Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One*, 6:e18565.

Gurgui, C. & Piel, J. (2010). Metagenomic approaches to identify and isolate bioactive natural products from

microbiota of marine sponges. *Methods in Molecular Biology*, *Springer Science + Business Media LLC*.

Haefner, B. (2003). Drugs from the deep: Marine natural products as drug candidates. *Drug Discov. Today* 8;536-544.

Hay, M.E. & Fenical, W. (1996). Chemical ecology and marine biodiversity: insights and products from the sea. *Oceanography,* 9(1);10-20.

Henstchel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J. & Moore, B.S. (2002). Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.,* 68(9);4431-4440.

Hentschel, U., Usher, K.M., & Taylor, M.W. (2006). Marine sponges as microbial fermenters. *FEMS Microbiol. Ecol.,* 55(2);167-177.

Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.*, 48;4688-4716.

Hirata, Y., & Uemura, D. (1986). Halichondrins – antitumor polyether macrolides from a marine sponge. *Pure & Appl. Chem.,* 58(5);701-710.

Hochmuth, T., Niederkrüger, H., Gernert, C., Siegl, A., Taudien, S., Platzer, M., Crews, P., Hentschel, U., & Piel, J. (2010). Linking chemical and microbial diversity in marine sponges: possible role for poribacteria as producers of methyl-branched fatty acids. *ChemBioChem,* 11;2572–2578.

Hrvatin, S., & Piel, J. (2007). Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *J. Microbiol. Meth.* 68;434-436.

Hu, G.-P., Yuan, J., Sun, L., She, Z.-G., Wu, J.-H., Lan, X.-J., Zhu, X. , Lin, Y.-C., Chen, S.-P. (2011). Statistical research on marine natural products based on data obtained between 1985 and 2008. *Mar. Drugs*, 9;514–525.

Hugenholtz, P., Goebel, B.M., & Pace, N.R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.,* 180(18);4765-4774.

Hunt, B., & Vincent, A.C.J. (2006). Scale and Sustainability of Marine Bioprospecting for Pharmaceuticals. *AMBIO,* 35(2);57-64

Ishikawa, J., & Hotta, K. (2006). FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol. Lett.*, 174(2);251-253.

Iwamoto, M., Shimizu, H., Muramatsu, I., & Oiki, S. (2010). A cytotoxic peptide from a marine sponge exhibits ion channel activity through vectorial-insertion into the membrane. *FEBS Lett.,* 584(18);3995-3999.

Jenner, M., Frank, S., Kampa, A., Kohlhaas, C., Pöplau, P., Briggs, G.S., Piel, J., & Oldham, N. J. (2013). Substrate specificity in ketosynthase domains from *trans*-AT polyketide synthases. *Angew. Chem., Int. Ed.,* 52(4);1143–1147

Kaufman, B., Richards, S., & Dierig, D.A. (1999). DNA isolation method for high polysaccharide *Lesquerella* species. *Indust. Crops Prod.*, 9;111-114.

Kennedy, J., Marchesi, J.R., & Dobson, A.D.W. (2007). Metagenomic Approach to exploit the Biotechnological Potential of the Microbial Consortia of Marine Sponges. *Appl. Microbiol. Biotechnol*. DOI 10.1007/s00253-007-0875-2.

Kim, U.J., Shizuya, H., de Jong, P.J., Birren, B., & Simon, M.I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucl. Acids Res.,* 20(5);1083-1085 PMCID: PMC312094

Kim, J.H., Feng, Z., Bauer, J.D., Kallifidas, D., Calle, P.Y., & Brady, S.F. (2010). Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers* 93;833-844.

Kitagawa, I., Kobayashi, M., Katori, T., Yamashita, M., Tanaka, J., Doi, M., & Ishida, T. (1990). Absolute stereostructure of swinholide A, a potent cytotoxic macrolide from the Okinawan marine sponge *Theonella swinhoei. J. Am .Chem. Soc.* 112(9);3710-3712.

Kleinkauf, H., & von Döhren, H. (1990). Antiobiotics - cloning of biosynthetic pathways. *FEBS Lett.,* 268(2);405-407.

Kohlhaas, C., Jenner, M., Kampa, A., Briggs, G.S., Afonso, J.P., Piel, J., & Oldham, N.J. (2013). Amino acid-accepting ketosynthase domain from a *trans*-AT polyketide synthase exhibits high selectivity for predicted intermediates. *Chem. Sci.,* 4;3212-3217.

Leal, M.C., Puga, J., Serôdio, J., Gomes, N.C.M., & Calado, R. (2012). Trends in the discovery of new marine natural products from invertebrates over the last two decades – where and what are bioprospecting. *PloS ONE,* 7(1);1-15.

Li, M.H.T., Ung, P.M.U., Zajkowski, J., Garneau-Tsodikova, S., & Sherman, D.H. (2009). Automated genome mining for natural products. *BMC Bioinformatics* 10;185-195.

Llewellyn, N.M., & Spencer, J.B. (2007). Enzymes line up for assembly. *Nature* 448(16):755-756.

Matsunaga, S., Fusetani, N., Hashimoto, K., & Walchli, M. (1989). Theonellamide F. a novel antifungal bicyclic peptide from a marine sponge *Theonella* sp. *J. Am. Chem. Soc.*, 111(7);2582-2588.

Martín, J.F. (2000). á-aminoadipyl-cysteinyl-valine synthetases in â-lactam producing organisms. From Abraham's discoveries to novel concepts of non-ribosomal peptide synthesis. *J. Antibiot*. 53. 1008e1021.

McGinnis, S., & Madden, T.L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucl. Acids Res*., 32;W20-W25.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucl. Acids Res.,* 39;W339–W346.

Mohamed, N.M., Enticknap, J.J., Lohr, J.E., McIntosh, S.M., & Hill, R.T. (2008). Changes in bacterial communities of the marine sponge *Mycale laxissima* on transfer into aquaculture. *Appl. Environ. Microbial.,* 74(4);1209-1222.

Munro, M. H., Blunt, J.W., Dumdei, E.J., Hickford, S.J., Lill, R.E., Li, S., Battershill, C.N., & Duckworth, A.R. (1999). The discovery and development of marine compounds with pharmaceutical potential. *J. Biotechnol., 70*;15-25.

Namikoshi, M., Rinehart, K.L., Sakai, R., Stotts, R.R., Dahlem, A.M., Beasley, V.R., Carmichael, W.W., & Evans, W.R. (1992). Identification of 12 hepatotoxins from a homer lake bloom of the cyanobacteria *Microcystis aeruginosa*, *microcystis viridis*, and *mycrocystis wesenbergii*: nine new mycrocystins. *J. Org. Chem., 57* (3);866–872.

Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., & Piel, J. (2008). Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol., 26*(2);225-233.

Ortega, V., & Cortés, J. (2012). Potential clinical applications of halicondrins in breast cancer and other neoplasms. *Breast Cancer: Targets Ther., 4*;9-19.

Phelan, V.V., Du, Y., McLean, J.A., & Bachmann, B.O. (2009). Adenylation enzyme characterization using ã⁻ $^{18}O_4$-ATP pyrophosphate exchange. *Chem. Biol., 16*;473–478.

Piel, J. (2004). Metabolites from symbiotic bacteria. *Nat. Prod. Rep., 21*;519–538.

Piel, J., Hui, D., Fusetani, N., & Matsunaga, S. (2004a). Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Env. Microbiol., 6*(9);921-927.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., & Matsunaga, S. (2004b). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei. Proc. Natl. Acad. Sci. USA, 101*;16222-16227.

Piel, J. (2006). Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals. *Curr. Med. Chem., 13*(1);39-50.

Piel, J. (2010). Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep., 27*;996-1047.

Prieto, C., García-Estrada, C., Lorenzana, D., & Martín, M.F. (2012). NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics 28*;426–427.

Proksch, P. (1994). Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs. *Toxicon 32*(6);639-655.

Proksch, P., Edrada, R.A., & R. Ebel (2002). Drugs from the Sea – Current Status and Microbiological Implications. *Appl. Micrbiol. Biotechnol., 59*;125-135.

Rawlings, B.J. (2001). Type I polyketide biosynthesis in bacteria (Part B). *Nat. Prod. Rep., 18*;231-281.

Ridley, C.P., Bergquist, P.R., Harper, M.K., Faulkner, D.J., Hooper, J.N.A., & Haygood, M.G. (2005). Speciation and biosynthetic variation in four dictyoceratid sponges and their cyanobacterial symbiont, *Oscillatoria spongeliae. Chem. Biol., 12*;397–406.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., & Kohlbacher, O. (2011). NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity. *Nucl. Acids Res., 39*;W362-W367.

Ruppert, E.E., Fox, R.S., & Barnes, R.D. (2004). Invertebrate zoology: a functional evolutionary approach. *Brooks/Cole-Thomson Learning, Inc. ISBN 0-03-025982-7 (989 p).*

Sadar, M.D., Williams, D.E., Mawji, N.R., Patrick, B.O., Wikanta, T., Chasanah, E., Irianto, H.E., van Soest, R., Anderson, R.J. (2008). Sintokamides A to E, chlorinated peptides from the sponge *Dysidea* sp. that inhibit transaction of the N-terminus of the androgen receptor in prostate cancer cells. *Org. Lett.* 10(21):4947-4950.

Sagar, S., Kaur, M., & Minneman, K.P. (2010). Antiviral lead compounds from marine sponges. *Mar. Drugs, 8*(10);2619-2638.

Sala, E., & Knowlton, N. (2006). Global marine biodiversity trends. *Annu. Rev. Environ. Resour., 31*;93–122.

Sambrook, J., & Russell, D.W. (2001). Molecular cloning: a laboratory manual, third edition. *Cold Spring Harbor Laboratory Press, New York.*

Scheuermayer M., Pimentel-Elardo, S., Fieseler, L., Grozdanov, L., & Hentschel, U. (2006). Microorganisms of sponges: phylogenetic diversity and biotechnological potential. *In Frontiers in Marine Biotechnology*, edited by Peter Proksch and Werner E.G. Müller. pp. 289-312.

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., de Long, E.F., Hutchinson, C.R., & Reid, R.C. (2005). Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta. Appl. Environ. Microbiol., 71*(8);4840-4849.

Schmidt, E.W., Obraztsova, A.Y., Davidson, S.K., Faulkner, D.J., Haygood, M.G. (2000). Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel ä-proteobacterium, "Candidatus *Entotheonella palauensis*". *Mar. Biol., 136*(6);969-977.

Schmidt, E.W., Nelson, J.T., Rasko, D.A., Sudek, S., Eisen, J.A., Haygood, M.G., & Ravel, J. (2005). Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella. Proc. Natl. Acad. Sci. USA, 102*(20);7315-7320.

Schweizer, E., & Hofmann, J. (2004). Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol. Mol. Biol. Rev., 68*;501-517.

Schwarzer, D., & Marahiel, M.A. (2001). Multimodular biocatalysts for natural product assembly. *Naturwissenschaften 88*;93-101.

Schwarzer, D., Finking, R., & Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. *Nat. Prod. Rep., 20*;275-287.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., & Simon, M. (1992). Cloning and stable

maintance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* 89;8794–8797

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., & Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA,* 103;12115-12120.

Taylor, M.W., Radax, R., Steger, D., & Wagner, M. (2007). Sponge-associated microorganisms: evolution, ecological, and biotechnological potential. *Microbiol. Mol. Biol. Rev.*, 71(2);295-347.

Teta, R., Gurgui, M., Helfrich, E.J.N., Künne, S., Schneider, A., van Echten-Deckert, G., Mangoni, A., & Piel, J. (2010). Genome mining reveals *trans*-AT polyketide synthase directed antibiotic biosynthesis in the bacterial phylum *Bacteroidetes*. *ChemBioChem,* 11;2506-2512.

Thakur, N.L., & Müller, W.E.G. (2004). Biotechnological potential of marine sponges. *Curr. Sci.* 86;1506-1512.

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., & Neilan, B.A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806*:* an integrated peptide–polyketide synthetase system. *Chem. Biol.*, 7(10);753-764.

Tsai, S.-C., & Ames, B.D. (2009). Structural enzymology of polyketide synthases. *Methods Enzymol.,* 459;17-47.

Turgay, K., Krause, M., & Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol,.* 6;2743-2744.

Turgay, K., & Marahiel, M.A. (1994). A general approach for identifying and cloning peptide synthetase genes. *Pept. Res.,* 7(5);238-241.

Uemura, D., Takahashi, K., Yamamoto, T., Katayama, C., Tanaka, J., Okumura, Y., & Y. Hirata, (1985). Norhalichondrin A: an antitumor polyether macrolide from a marine sponge. *J. Am. Chem. Soc.*, 107(16);4796-4798.

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., & Piel, J. (2015). Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nature Chem. Biol.* 11:705-712.

Unson, M.D., & Faulkner, D.J. (1993). Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia,* 49;349-353.

Unson, M.D., Holland, N.D., & Faulkner, D.J. (1994). A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar. Biol.* 119;1-11.

Uria, A.R. (2012). Investigating Natural Product Biosynthesis in Uncultivated Symbiotic Bacteria of the Marine Sponge *Theonella swinhoei*. PhD Dissertation, University of Bonn, Germany. (218 p).

Uria, A.R., Munifah, I., & Chasanah, E. (2007). Isolation and enzyme fragmentation of genomic DNA from uncultured microbial symbionts of *Theonalla* sp. sponge as prerequisite for metagenomic library construction. *Ind. Fish. Res. J.,* 13(2);125-130.

Uria, A.R., & Piel, J. (2009). Cultivation-independent approaches to investigate the chemistry of marine symbiotic bacteria. *Phytochem. Rev.,* 8;401–414.

Uria, A.R., Fawzya, N.Y., & Chasanah, E. (2005). Novel molecular methods for discovery and engineering of biocatalysts from uncultured marine microorganisms. *J. Coast. Develop.,* 8(2);49-74.

Velasco, A.M., Leguina, J.I., & Lazcano, A. (2002). Molecular evolution of the lysin biosynthetic pathways. *J. Mol. Evol.,* 55;445-459.

Wakimoto T., Egami, Y., Nakashima, Y., Wakimoto, Y., Mori, T., Awakawa, T., Ito, T., Kenmoku, H., Asakawa, Y., Piel, J., & Abe, I. (2014). Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat. Chem. Biol.,* 10;648–655.

Webster N.S, & Hill, R.T. (2001). The culturable microbial community of the Great Barrier Reef sponge *Rhopaloeides odorabile* is dominated by an á-Proteobacterium. *Mar. Biol.* (Berl), 138;843–851.

Westenberg, M., Bamps, S., Soedling, H., Hope, I.A., & Dolphin, C.T. (2010). *Escherichia coli* MW005: lambda Red-mediated recombineering and copy-number induction of *oriV*-equipped constructs in a single host. *BMC Biotechnol.* 10(27);1-8.

Wild, J., Hradecna, Z., & Szybalski, W. (2002). Conditionally amplifiable BACs: switching from single copy to hig-copy vectors and genomic clones. *Genome Res.,* 12;1434-1444.

Wilson, M.C., Mori, T., Rückert, Ch., Uria, A.R., Helf, M.J., Takada, K., Gernert, Ch., 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., & Piel, J. (2014). *An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature,* 506(7486);58-62.

Yadav G., Gokhale, R.S., & Mohanty, D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.*, 328;335–363.

Youssef, T.A., & Mooberry, S.L. (2006). Hurghadolide A and swinholide I, potent actin-microfilament disrupters from the Red Sea sponge *Theonella swinhoei. J. Nat. Prod.* 69(1);154-157.

Zehner, S., Kotzsch, A., Bister, B., Süssmuth, R.D., Méndez, C., Salas, J.A., van Pée, K-H. (2005). A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* 12(4);445–452.

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., & Jensen, P.R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE,* 7(3):e34064.