# New Methods for Medicinal Chemistry – Universal Gene Cloning and Expression Systems for Production of Marine Bioactive Metabolites

Walter C. Dunlap[1], Marcel Jaspars[2], Daslav Hranueli[3], Christopher N. Battershill[1], Nataša Perić-Concha[4], Jurica Zucko[3], Stephen H. Wright[2] and Paul F. Long[4,*]

[1]*Australian Institute of Marine Sciences, Townsville, Queensland, Australia*

[2]*Department of Chemistry, University of Aberdeen, Scotland, UK*

[3]*Faculty of Food Technology and Biotechnology, University of Zagreb, Croatia*

[4]*School of Pharmacy, University of London, UK*

**Abstract:** Natural products from symbiotic or commensal associations between marine invertebrate and microbial organisms show exceptional promise as pharmaceuticals in many therapeutic areas. An economic and sustainable global market supply due to difficulty of synthesis is cited as the main obstacle for exploitation of these otherwise exciting marine bioactive compounds [1]. Different strategies have been evoked to overcome this impediment as long-term harvesting of wild stocks from the environment is considered unsound, and other modes of production based on biosynthesis, such as aquaculture, have not yet been proven as reliable [2]. One option is to clone the genes encoding the biosynthetic expression of a lead metabolite into a surrogate host suitable for industrial-scale fermentation. To facilitate this goal we are developing a universal system to clone and express genes responsible for biosynthesis of natural products from both eukaryotic and prokaryotic partners of marine symbioses. The ability to harness the complete meta-transcriptome of entire biosynthetic pathways is particularly valuable where the biogenesis of a target natural product occurring within a complex symbiotic association is unclear.

**Keywords**: Marine science, microbial genetics, marine natural products, drug development, heterologous expression, metagenomic cloning, bioinformatics, genomics screening.

## INTRODUCTION

The rich structural and stereochemical attributes of natural products [3] promote these compounds as valuable molecular scaffolds to explore their chemotherapeutic potential. Natural products have historically made an impressive contribution as new lead candidates in drug discovery programs, and a small number are now advancing well through clinical trial [4,5]. The World's seas and oceans cover over 70% of our planet, yet these ecosystems have been largely unexploited as a source of novel compounds by the pharmaceutical industry. Discovery programs to date have been limited to physiological diving depths, and sampling of deeper ocean habitats is limited to utilizing grabs and dredges with limited capture capacity. Until very recently, tools for collecting and maintaining the vast uncultured resources of marine microbial communities have not been well engaged in biodiscovery activities. There is now unprecedented opportunity to sample the globes oldest and most diverse genomes enabling new strategies to rapidly progress development of these unexplored chemical spaces that would add considerable value to pharmaceutical lead generation efforts. This is especially welcome in a post-genomic era where disease indications continue to expand rapidly and the need for remedy is vital. With an ability to culture less than 1% of the predicted microbes found in soils, it is remarkable that these terrestrial organisms have been such a rich source of biologically active natural products for over 50 years. Access to small molecules made by the viable yet uncultivated majority has now been made possible by the isolation of metagenomic DNA and heterologous expression of biosynthetic pathways in a fermentable host. Herein we review strategies to translate and adapt this technology for the sustainable supply of bioactive natural products obtained from marine environments.

### Marine Natural Products: Potential and Applications

It is generally accepted that life evolved billions of years ago in the primitive sea, predominantly in the energy-rich, shallow photic layers. It is the modern oceans that contain the greatest diversity of life on Earth. In many extant phyla ancient biosynthetic pathways are conserved through evolutionary constrains determined by the prevalence of asexual reproduction. Marine micro- and macro-organisms are a rich source of chemically diverse metabolites with a wide range of potential use in commercial and medicinal applications. In particular, secondary metabolites of marine invertebrates show exceptional promise as potential pharmaceuticals in diverse therapeutic areas in treating cancer, pain, inflammation and viral infections [4-6]. Most marine invertebrates harbour microorganisms that include bacteria, cyanobacteria, fungi and eukaryotic algae within host tissues where they reside as extra- and intra-cellular symbionts. In some sponges microbial symbionts may

*Address correspondence to this author at the School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, UK; Tel: +44-(0)20-77535868; Fax: +44-(0)20-77535868; Email: paul.long@pharmacy.ac.uk

**0929-8673/06 $50.00+.00** 

constitute as much as 40% of the volume of the holobiont [7] and can be accommodated within specialist host sponge cells, the bacteriocytes [8] Fig. (**1**). Many of these invertebrates are microplankton filter feeders allowing the assimilation of consumed microorganisms from inhaled seawater by phagocytosis. The environmental, metabolic and biosynthetic relationships between host and endobionts are usually complex and remain poorly understood. Numerous products isolated from marine invertebrates are structurally homologous to known metabolites of strict microbial origin, suggesting that commensal microbiota contribute substantially to the biosynthesis of host metabolites. Molecular methods have recently been applied to fully characterize the phylogenetic diversity and host specificity of microbial associations in marine sponges [9-13], and available evidence reveals that certain sponge-specific symbionts, most of which (>98%) have not yielded to culture, may have co-evolved with their hosts [14,15]. Sponges (phylum Porifera) are phylogenetically the oldest and simplest multicellular animals of the extant basal Metazoa (ca. 800 million years), and as microbial filter feeders they are abundantly exposed to potential infection by trophic microorganisms. Accordingly, the unique specificity of sponge-microbial associations offers an ideal situation to examine the molecular functions of host-cell specificity in the primitive innate immune systems of early evolution (http://www.pirweb.org/pir04b_marine.htm).

Several metabolites from marine invertebrates that are in preclinical or clinical trial phases [4-6], such as the sponge anti-mitotic agents (+)-discodermolide 9 (**1**) from the sponge *Discodermia dissoluta* [16] and halichondrin B (**2**) from the sponge *Lissodendoryx n. sp.* [4], are alleged to be products not of the invertebrate animal, but derived from their microbiotic consortia [17]. In the case of the anti-tumour active bryostatin 1 (**3**) from the marine bryozoan *Bugula neritina*, molecular evidence favours a microbial origin of these macrocyclic lactones produced by a characterised, but yet uncultured γ-protobacterial endobiont [18-21]. In contrast, symbiotic actinomycetes from the marine tunicate *Aplidium lenticulum* have been cultured to yield a

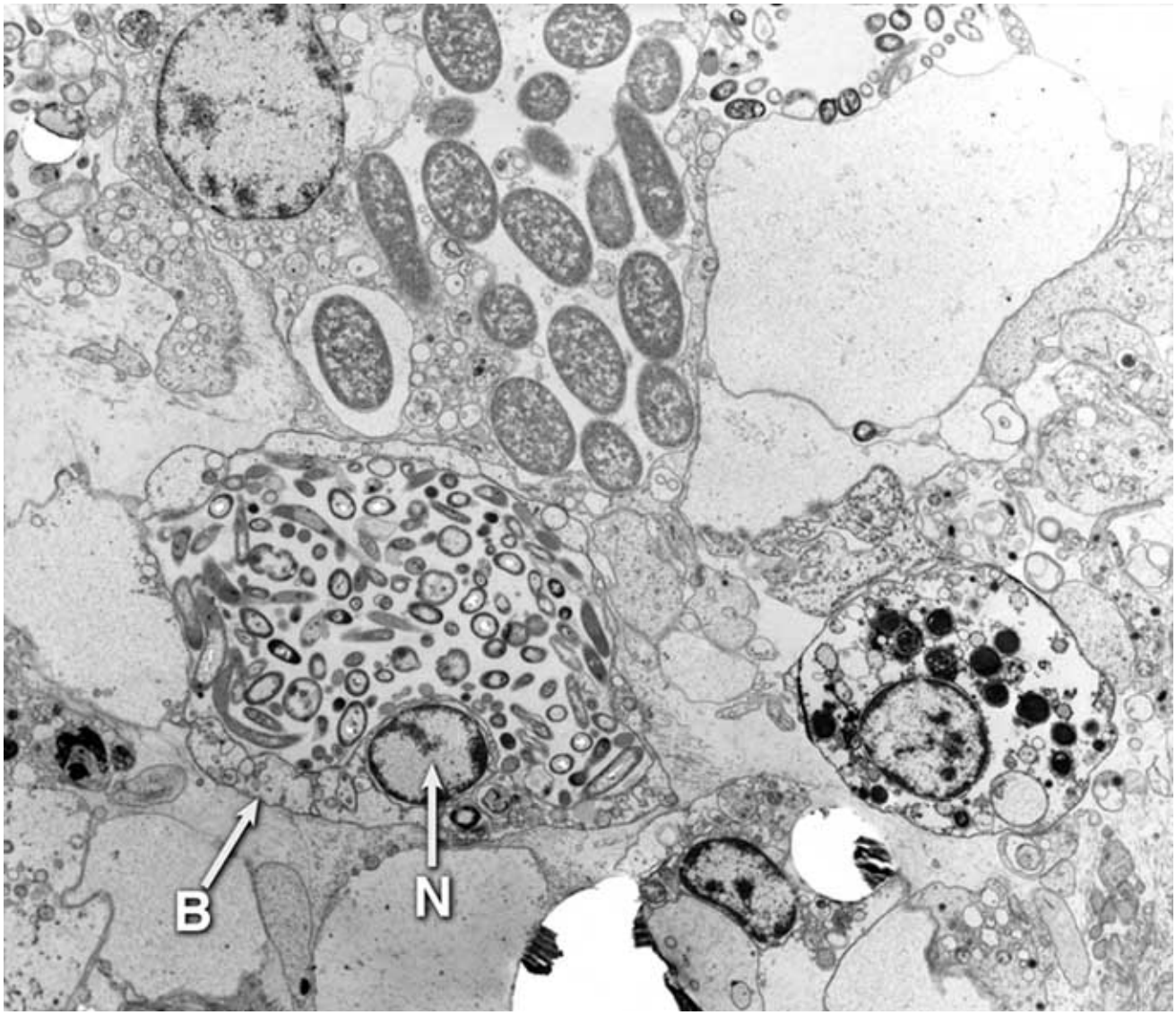
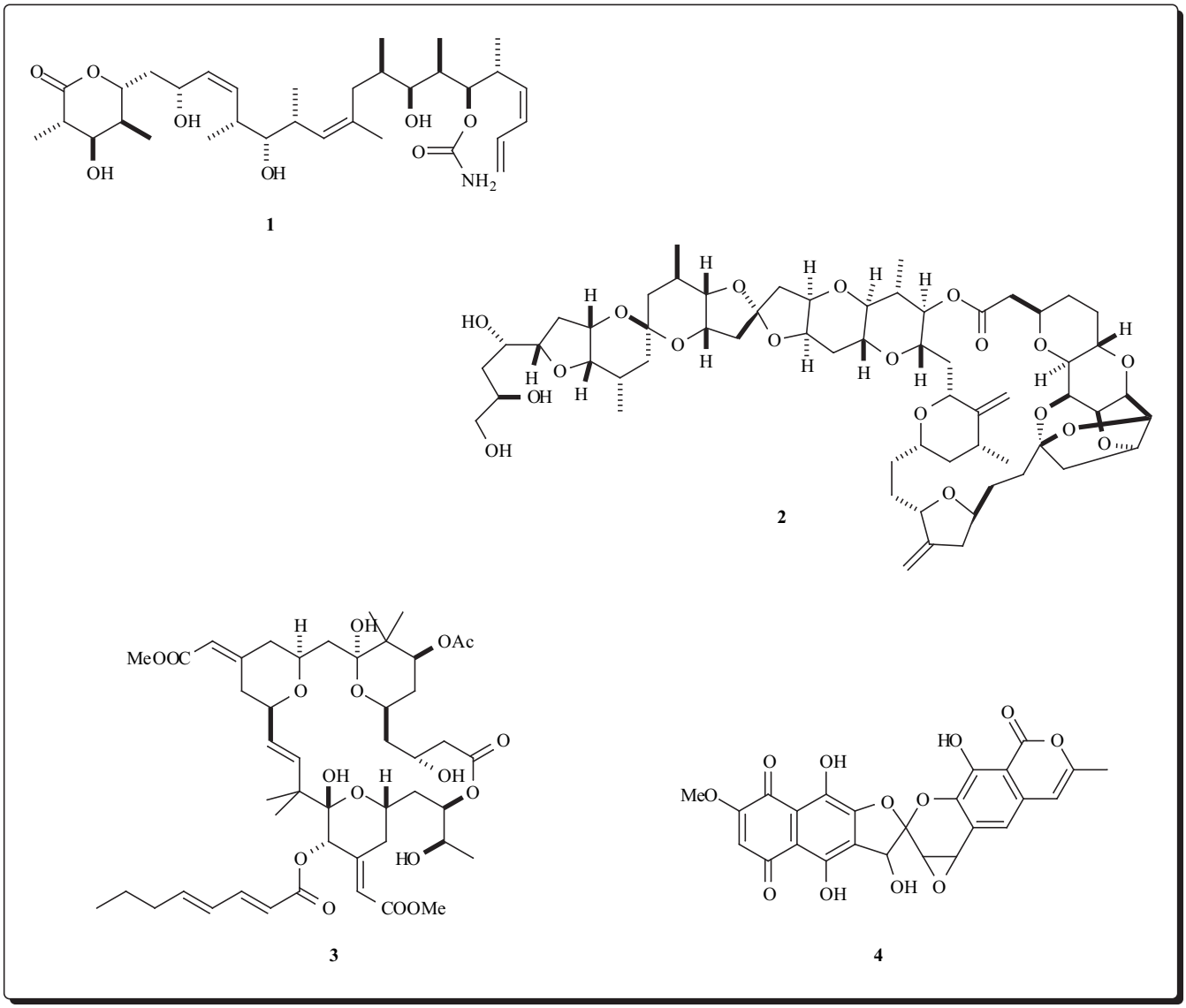


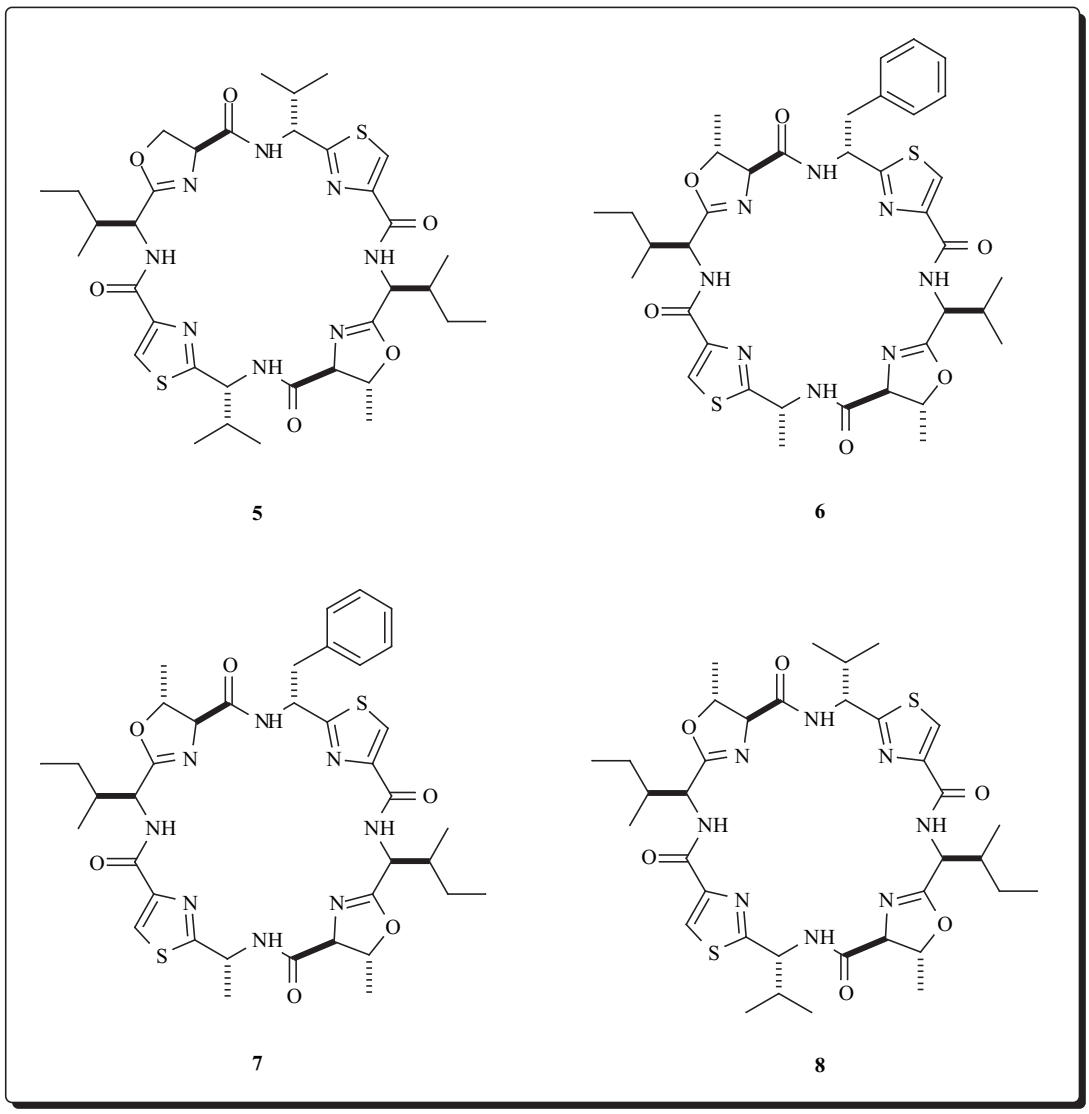

**Fig. (1).** Transmission electron micrograph (TEM) of a sectioned tissue from the New Zealand sponge *Mycale hentscheli* showing a bacteriocyte (**B**) containing a nucleus (**N**) surrounded by a large population of intracellular bacteria.



*Streptomyces sp.* isolate (strain JP95) that produces the telomerase inhibitor griseorhodin A (**4**), from which the biosynthetic gene cluster has been sequenced [22].

The patellamides, such as patellamide A (**5**), patellamide C (**6**), patellamide D (**7**) and ascidiacyclamide (**8**) [23,24], are a highly conserved family of oxazoline and thiazole containing cyclo-octapeptides from the tropical Indo-Pacific aplousobranch ascidian (sea-squirt) *Lissoclinum patella* [25,26]. These compounds have differential cytotoxic properties targeting fibroblast and tumour cell lines [23]. Evidence for nonribosomal peptide synthetase (NRPS) biosynthetic genes within *Prochloron* species symbiotic to the Didemnid family of ascidians would appear to favour the prokaryote symbiont as the origin of these cyclic peptides [26]. Sequencing the genome of *Prochloron* symbionts from a Palau specimen of *L. patella*, however, has revealed that the pathway is ribosomal [27]. Patellamide D, possibly as a binuclear copper complex [28], has additional pharmacological significance as a selective antagonist for reversing the multidrug resistant CEM/VLB100 human leukemic cell line towards vinblastine, colchicine and adriamycin treatment [29]. Potential clinical development of this pharmacophore is likely to be encouraged by achieving a sustainable supply of these bioactive marine metabolites.

A vast array of natural products of both marine and terrestrial origin are synthesised by large multifunctional modular polyketide synthases (PKS), NRPSs, and hybrid NRPS-PKS enzymes organized as a co-linear arrangement of catalytic centers grouped into modules, with each module responsible for the sequential catalysis of one complete cycle of polyketide or polypeptide chain elongation [30-33]. For example, polyketide metabolites are assembled through a common mechanism by the condensation of small carboxylic acids. Structural differences in polyketides stem mostly from variations in the number of condensations, the type of starter and extender units incorporated, and the extent of stereo-specific reduction occurring during each cycle of ketide extension [34]. The majority of polyketides made by microbiota are from modular PKSs, which direct polyketide synthesis on a protein template whereby each module is responsible for selecting, incorporating and processing the appropriate carboxylate unit. The architecture of modular



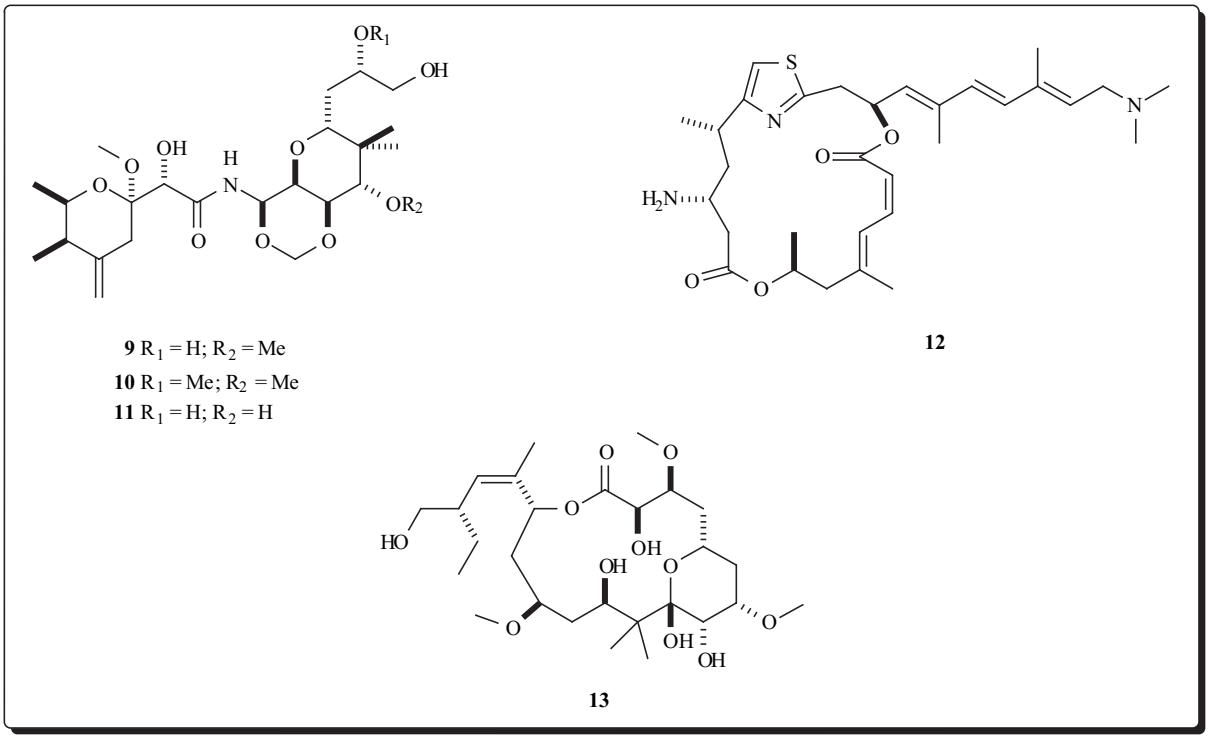
**5** **6**

**7** **8**

PKS (and NRPS) enzymes allows genetic reprogramming to produce custom designed metabolites [35,36]. The availability of a growing number of modular PKS enzymes of defined sequence has led to the successful production of many novel and structurally diverse polyketides by genetic manipulation of PKS pathways, albeit at low titres [37,38].

The New Zealand sponge *Mycale hentscheli* (Demospongidiae: Poecilosclerida: Mycalidae) offers three separate classes of PKS compounds with pharmaceutical potential. Mycalamides (**9** and **10**) were first described as antiviral agents [39] and together with mycalamide D (**11**) [40] were characterised as eukaryotic cytotoxins that inhibit protein synthesis causing apoptosis [41,42]. Pateamine (**12**) having immuno-suppressive and apoptotic properties [42,43] has been isolated in discrete populations of *M. hentscheli* that lack mycalamides [44]. Peloruside A (**13**) was additionally isolated from *M. hentscheli* as a cytotoxin structurally unrelated to pateamine or the mycalamides [45]. Peloruside A is a 16-membered polyketide macrolide similar to epothilone having potent antimitotic activity with paclitaxel-like microtubule-stabilizing properties that arrest cells in the $G_2$/M phase of the cell cycle [46]. The bioactive metabolite composition of *Mycale hentscheli* collected from New Zealand is remarkably site specific, and intra-specific concentrations are variable within regional sites [47,48]. Such chemotypic variation of metabolite expression by *Mycale* conspecifics argues strongly that the sponge residing under different abiotic conditions sustain divergent populations of bacteriocyte endobionts, possibly clonal variants of host-specific cyanobacteria [49], that are essential to the biosynthesis of these active secondary metabolites.

Even more structurally diverse and abundant than PKS and NRPS metabolites are the isoprenoids that have numerous functional roles in primary and secondary metabolism, of which terpenoids are a class particularly abundant in plants. Over 30 cDNAs encoding plant monoterpene (C10), sesterterpenes (C15) and diterpene (C20) synthases have been cloned and characterised [50]. Unlike plant terpenes, however, very little is known of the biogenesis and genomic organisation of terpenes in symbiotic marine organsims, including sponge sestertepenes (C25) such as the manoalides. One study has revealed that sesterterpene biosynthesis is species-specific in diatoms utilising either the mevalonate (MVA, as in higher plants) or methylerythritol (MEP) pathways [51], which must be a



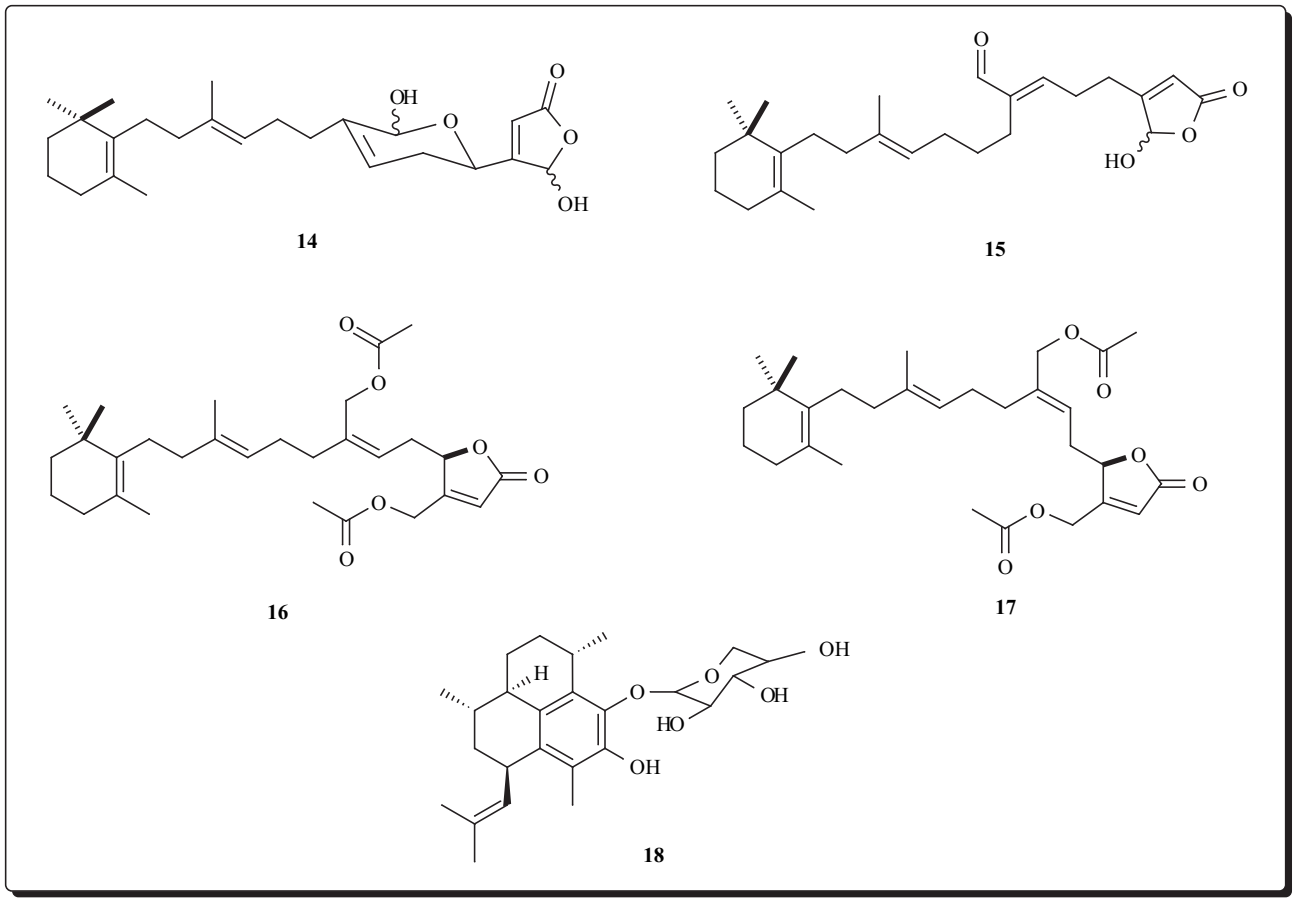

**9** $R_1$ = H; $R_2$ = Me
**10** $R_1$ = Me; $R_2$ = Me
**11** $R_1$ = H; $R_2$ = H

**12**

**13**

consideration when undertaking heterologous expression of terpenoid products [52].

The sesterterpenes, manoalide (**14**), seco-manoalide (**15**), *E*-neomanoalide (**16**) and *Z*-neomanoalide (**17**) were first reported from the Palauan sponge, *Luffariella variabilis*, as antibacterial agents [53,54]. Manoalide was subsequently found to have potent anti-inflammatory properties and to irreversibly inhibit phospholipase $A_2$ ($PLA_2$) [55-59].




**14**

**15**

**16**

**17**

**18**

Manoalide has also been shown to be a potent inhibitor of cellular calcium mobilization [60]. A large number of analogues have since been isolated from a variety of *Luffariella* species having similar properties [61-68]. Manoalide was progressed through Phase I clinical trials for topical treatment of psoriasis [69], but therapeutic quantities would not diffuse through the skin and further testing was suspended. Manoalide is used extensively as a pharmacological probe for $PLA_2$ and is available commercially in small quantities.

Natural manoalide exists as a mixture of diastereoisomers due to the presence of two hemiacetal moieties at C24 and C25. Although numerous syntheses of manoalide have been reported [70-74] including enantiospecific syntheses [75,76], low synthetic yields are not viable to sustain a large-scale market supply. Other than the elucidation of the anti-inflammatory, diterpene glucoside pseudopterosin (**18**) [77-79] produced by symbiotic *Symbiodinium sp*. dinoflagellates of the sea whip (gorgonian) *Peusdopterogordia elisabethae* [80], little is known about the biosynthetic construction of terpenes in marine invertebrates. Successful cloning of terpene synthase (TPS) genes encoding the presumed biosynthetic pathway for manoalide Fig. (**2**) underway in our laboratory may reveal much about the biogenic origin and genomic architecture of sesterterpene biosynthesis in *Luffariella*, and the TPS pathway of marine organisms in general.

## ACHIEVING A SUSTAINABLE SUPPLY

The foregoing examples touch lightly on the established wealth of marine secondary metabolites available for pharmaceutical advancement and hint at the vast potential of those yet to be discovered. The quest for modern medicine to improve health in aging populations has never been better poised for the development of "drugs from the sea". Yet, the main obstacle to exploiting the chemical diversity of marine bioactive metabolites for pharmaceutical investment has been the economic issue of providing a sustainable and robust supply [1]. Different strategies have been attempted to overcome this impediment, as long-term harvesting of wild stocks from the environment is generally unsound. Aquaculture, tissue culture, symbiont culture and chemical synthesis are possible but seldom provide viable and

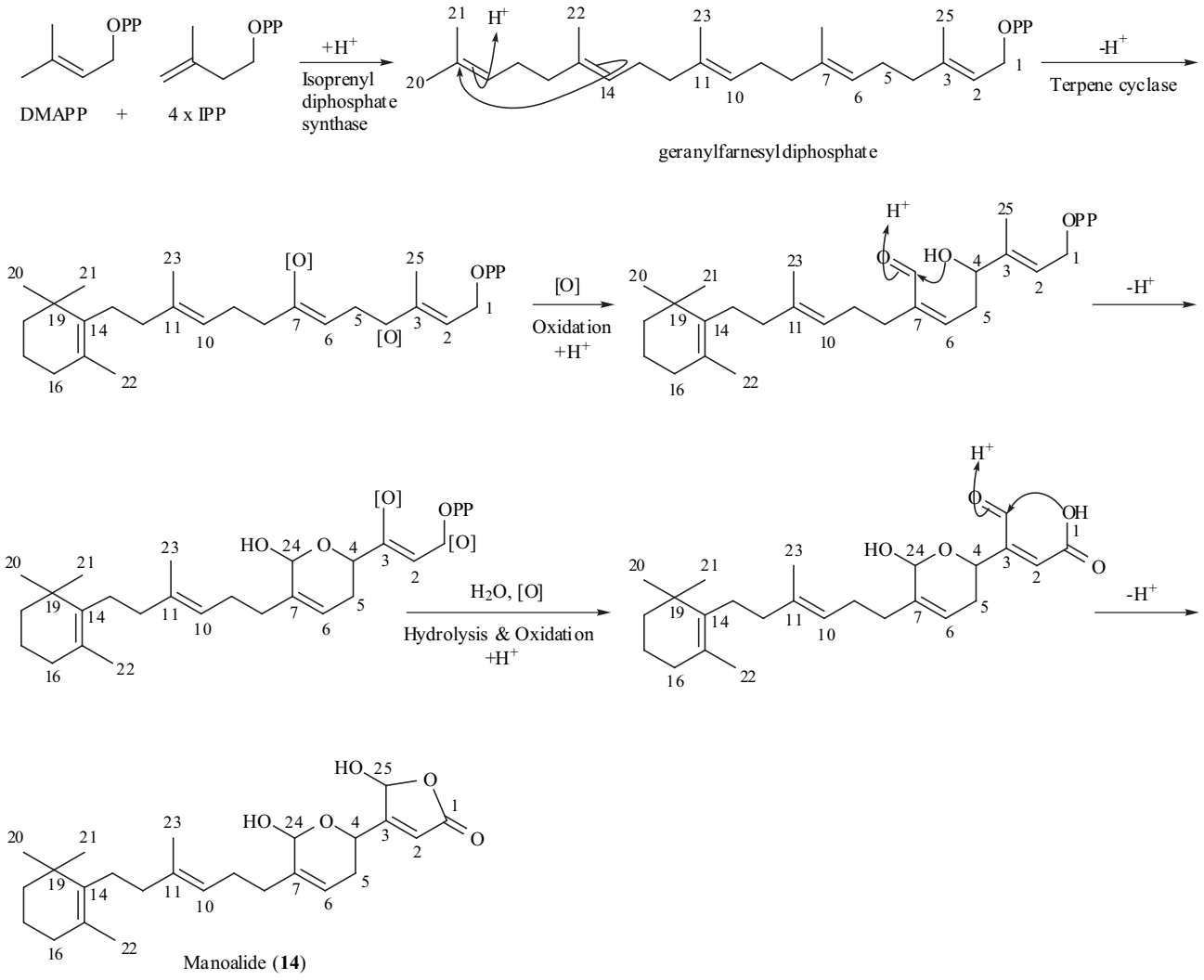
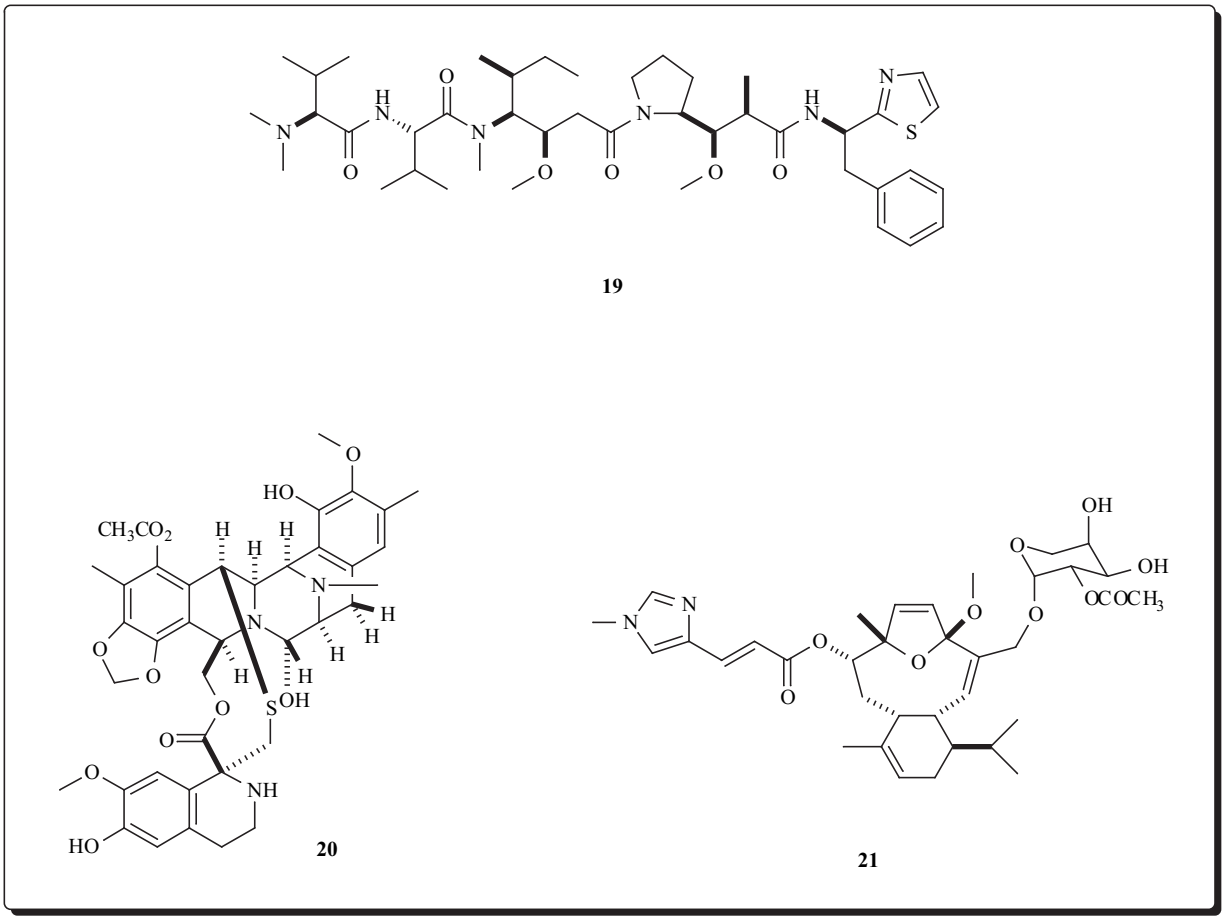



**Fig. (2).** Proposed pathway for the biosynthesis of manoalide (**14**); DMAPP = γ,γ-dimethylallyl pyrophosphate; IPP + isopentenyl diphosphate.

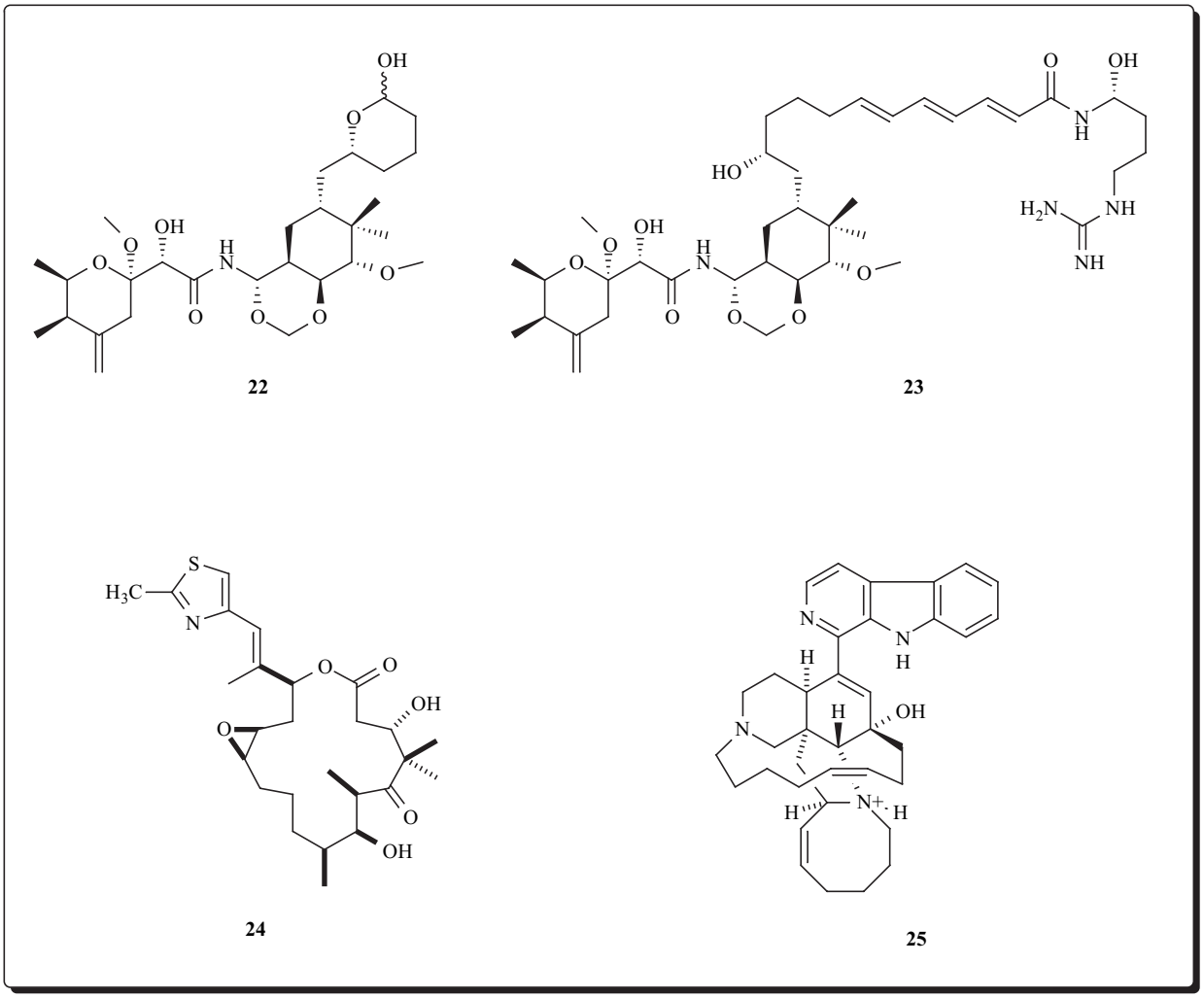


economic yields to supply a global market demand [2,81]. Pharmacologically active marine metabolites are often produced at exceedingly small concentrations in species with low distributions or occur in populations that are highly site-specific, so that obtaining sufficient quantities by large-scale collection is ecologically unsustainable. Total or semi-synthesis has been the traditional choice of supply, but the molecular complexity of many candidate molecules require multi-step processes having inefficient yields that prove too costly for commercial-scale production. Chemical synthesis has been useful to confirm structural elucidation and to prepare small quantities for clinical investigations, notable examples being (+)-discodermolide (**1**) [82] with a plan towards commercial synthesis (83), the mitotic inhibitor dolastatin 10 (**19**) derived from the Indian Ocean sea-hare *Dolabella auricularia* [84] and ecteinascidin 743 (**20**) from the tunicate (sea-squirt) *Ecteinascidia turbinata* [85] including a semi-synthetic preparation starting from the bacterial cyanosafracin B antibiotic available in kilogram quantities from industrial-scale fermentation of *Pseudomonas fluorescens* [86].

Mariculture is a proven method for the supply of several important marine natural products such as halichondrin B (**2**) from the marine sponge *Lissodendoryx n sp.* [87,88], bryostatin 1 (**3**) from the bryozoan *Bugula neritina* [89,90], ecteinascidin 743 (**20**) from the Caribbean tunicate *Ecteinascidia turbinate* [80,90] and the diterpene glycoside eleutherobin (**21**) from the gorgonian *Erytheopodium caribaeorum* [91]. Yields achieved by mariculture in these examples are typically 5-12 x $10^{-4}$ % of the wet weight biomass, although some evidence is available that yields can be enhanced at harvest by culture manipulation [87]. Of particular concern is the reliability of production by in-sea culturing that has been found vulnerable to destruction by climatic and environmental events [90] combining additional exposure to devastation by infection or pests.

Microbial symbionts have been implicated in the biosynthesis of many bioactive metabolites [21] as elegantly demonstrated for theopederin A (**22**) and onnamide A (**23**) TPS biosynthesis in the sponge *Theonella swinhoei* [92]. However, less than 1% of the commensal microbiotic consortia are amenable to laboratory culture [93]. Even if microbial populations can be successfully separated, undefined metabolic factors from the sponge are required for the symbionts to grow *ex hospite*. A successful exception is the telomere-inhibiting griseorhodin A (**4**) [94] cultured from a symbiotic *Streptomyces sp.* and the anti-tumour alkaloid manzamine A (**24**) obtained from the bacterium *Micromonospora sp.* isolated from a deep-water Indonesian sponge [95,96]. While tissue culturing of host primorphs may hold promise [97,98], the utility of enabling methodologies has not been realised [99].

The use of molecular biology to clone the genes responsible for the biosynthetic assembly of a desired marine metabolite for heterologous expression in a surrogate host offers the greatest promise to achieve economical, commercial-scale production that is both sustainable and reliable. Although genes encoding the biosynthesis of




griseorhodin A (**4**) [22] and several other marine metabolites have been entirely sequenced [100-102], heterologous expression has been elusive despite considerable efforts [20]. While biosynthetic expression of several bioactives including the potent anti-mitotic epothilones (**25**) [103,104] derived from terrestrial sources have been accomplished, the first surrogate expression of a marine metabolite was achieved from an Australian specimen of *Lissoclinum patella* yielding the cyclic octapeptides, patellamide D (**7**) and ascidiacyclamide (**8**) [105]. The challenge remains, however, to identify and clone metagenomic DNA from host tissues and the uncultured microbiotic consortia of a complex marine invertebrate (sponge) for heterologous expression of bioactive candidates for pharmaceutical development.

## Molecular Genetic Approaches Towards Sustainable Supply

Fuelled by the realization that genetic engineering has the potential to produce structural analogues of bioactive compounds by combinatorial biosynthesis [106], there continues to be considerable interest at the molecular level in studying how secondary metabolites are synthesized by cultivated terrestrial microbes [107]. Many of these compounds share structural similarities with bioactives isolated from marine sources [93]. Indeed, there is some debate as to whether actinomycetes cultured from the marine environment are derived from the run-off of terrestrial spores [108], although abundant evidence suggests that obligate marine species do occur [109,114] particularly those among the seawater-requiring taxon *Salinospora* gen. nov. [115,116], including a symbiotic *Salinospora* sp. of the marine sponge *Pseudoceratina clavata* [117]. Given that marine microbiota are a largely unexplored resource for drug discovery, it is logical that lessons learnt from investigating the molecular biology of terrestrial secondary metabolism are now being applied to marine systems [118]. Taking a terrestrial microbial genetics approach, heterologous pathway expression has become a realistic strategy to overcome the supply impediments facing the clinical development of marine bioactives.

The *Prochloron* sp. symbiont of *L. patella* was thought to be the likely source of the patellamide metabolites since their core structure is typical of a prokaryotic NRPS origin [26], even though studies have shown that these compounds are localized predominately in the host tunic [25]. Although *Prochloron* is recalcitrant to laboratory culture, this obligate symbiont can easily be separated in large quantities from the cloacal cavity of the holobiont. Long *et al.* [105] were able to 'shotgun' clone large random-sized fragments of genomic DNA from the *Prochloron* sp. symbiont into *Escherichia*

*coli* DH10B cells using pPAC-S1 and pPAC-S2 bacterial *E. coli – Streptomyces* artificial chromosome (BAC) shuttle vectors [119] to obtain the heterologous expression of patellamide D (**7**) and ascidiacyclamide (**8**). This provided conclusive evidence that *Prochloron* is the ultimate biosynthetic source of the patellamides. In an alternative approach, assuming a ribosomal origin for the patellamides, Schmidt *et al.* [27] used "reverse genetics" to correlate the amino acid composition of the patellamides A (**5**) and C (**6**) to the annotated putative patellamide biosynthetic genes from the *Prochloron didemni* sequencing project. In the absence of vigorous structural elucidation data, production of only patellamide A was reported in exeptionally low yield after heterologous expression of the genes, again using *E. coli* as the surrogate host. The disadvantages limiting future use of this genomic approach for heterologous expression of marine bioactives are clear. Assigning the true biosynthetic origin for any biosynthetic compound is problematic given the large diversity of microorganisms that live in close assemblage with many invertebrates. Even if such an assignment were possible, our current limitations in separating and culturing these symbiotic microbes makes obtaining sufficient quantities of homogenous genomic DNA for sequencing near impossible, assuming sequencing and annotation facilities capable of handling a metagenome project are also available.

The strategy of Long *et al.* [105] also has its limitations. A universal cloning technique suitable to 'shotgun' genetic material encoding biosynthetic pathways from prokaryote as well as eukaryote consortia of complex marine assemblages is required if the full potential of recombinant drugs from the sea is to be realized. "Metagenomics" has been described by Handelsman [120] as the study of heterogeneous genomic DNA derived from a defined source. Constructing genomic DNA libraries by cloning fragments from a pool of genomic DNA into a surrogate host has been used with great effect to study secondary metabolism in complex soil communities from viable yet uncultivated bacteria [121-127], and has been used likewise to examine the genomic structure of marine planktonic microbial assemblages [128,129]. Crucial is the judious choice of vector that can accommodate not only large fragments of DNA to capture entire biosynthetic pathways, but which can also be shuttled between various hosts increasing the likelihood that transcription and translation barriers are overcome successfully. There are gaps in our understanding of the community composition of *Prochloron* spp. at the genetic level, so we do not know whether the genomic DNA library of Long *et al.* [105] can truly be described as metagenomic. The strategy to construct this clone library with an *E. coli - Streptomyces* shuttle vector [130] was to allow expression and screening in a genetically amenable *E. coli* bacterium, but permits expressing the library in an alternative and physiologically more flexible *Streptomyces* host, if required. A sustainable supply of the bryostatin anticancer agents produced by the Gram-negative, marine bacterium *Endobugula sertula* has yet to be achieved for the uncultured symbiont of the bryozoan *Bugula neritina* [131]. Thus, applying the 'shotgun' cloning approach of Long *et al.* [105] to express bryostatins in a surrogate host would be an attractive proposition to attain sustainable production of these valuable bioactives.

Statistical factors also need to be taken into account when constructing a metagenomic library [132]. The size of the genome or combined genomes, the size of the inserts used and the predicted length of the biosynthetic gene cluster all need to be considered, as well as the mode of gene expression, which can be independent based on a transcriptional fusion or dependent on expression signals located on the vector. Only independent gene expression, such as was achieved for the patellamide [105], shows a reduction in clone library size with an increase in the size of the insert, whereas expression via transcriptional fusion shows an effective upper limit on insert size, and the dependent expression necessitates the use of very large libraries ($10^7$ clones) at all times. The use of a BAC based strategy with large inserts will accordingly be useful only in the case of independent gene expression providing a more labour intensive method, but with more specific medium- or low-throughput screening to be employed to deconvolute the clone library.

Marine metagenomic libraries to date have been limited to the expression of composite prokaryotic genomic DNA, although eukaryotic DNA by definition must also be represented in these libraries given the complex source material from which the DNA was extracted. Expression of eukaryotic pathways in a prokaryote host would appear, at first hand, to be a redundant strategy because of incompatibilities in post-transcriptional mRNA processing. However, eukaryotic expression systems do exist [133], and modification of these or prokaryotic surrogates into 'super-hosts' to bypass transcription, translation or biosynthetic barriers could become a reality in the future, for example, by the additional cloning of RNA processing enzymes [134]. Selection of an appropriate host is key to sustainable production on an industrial scale to sustain a future global supply. Providing as near a compatible cellular environment for transcription, translation and metabolite production as the native host would be a surreptitious consideration to guide the choice of a surrogate host. Transferring secondary-metabolite production between related strains of micro-organisms has been achieved for both terrestrial as well as marine metabolites, as for example production of the telomerase inhibitor griseorhodin A (**4**) in *Streptomyces lividans* [22]. Increasingly, transfer and functional expression of secondary-metabolite pathways between taxonomically distant organisms are being reported [103] with *E. coli* proving to be a productive host in now many examples [27,104,105,122,124,125]. Conversely, metabolic pathway engineering may be required to introduce precursor substrates not normally found in *E. coli* to support surrogate biosynthesis [135,136]. This highlights the need to develop effective genetic tools for the expression of DNA encoding secondary-metabolic pathways, most notable the need to shuttle between metabolically distinct hosts without the laborious and technically challenging need to sub-clone large fragments of DNA. Recently, a new bacterial artificial chromosome vector has been reported that can be manipulated in *E. coli* and then transferred by conjugation into the genome of two recipient organisms containing the *Streptomyces* phage C31 phage-attachment site *attB* for integration [137]. The source biosynthetic code most likely to be useful for capturing all possible consortia in a marine assemblage will come from interrogation of the (meta)-

transcriptome. Harnessing the transcriptome would overcome the presence of introns which are the main obstacle presently hampering 'shotgun' cloning and expression of eukaryote pathways, although biotic and abiotic conditions could also affect variations in metabolite production that limit the abundance of the required mRNA transcripts.

## BIOINFORMATICS FOR ANALYSIS OF METAGENOMIC LIBRARIES

The overwhelming majority of uncultivated microbes present in the environment, either free living or part of a complex symbiotic assemblage, provide a vast resource for the discovery of hitherto unknown novel compounds by the pharmaceutical industry [138]. Selecting individual clones of potential biological interest from metagenomic libraries has traditionally followed either functional or genetic-based screening approaches. In the functional approach, successfully pioneered by biotechnology companies such as Cubist Pharmaceuticals [http://www.cubist.com] and Diversa [http://www.diversa.com], genes from a metagenomic library are heterologously expressed in a surrogate host and then compound extracts screened against biological assays to detect the function of interest [139]. In the genetics-based approach, clones are selected based on the presence of either phylogenetic or metabolic genes [140,141]. Shotgun sequencing of entire metagenomic libraries has emerged as a third option, allowing us to capture genes that are either too diverged from currently known genes to be heterologously expressed in common hosts, or to be amplified using degenerate primers based on sequence alignments of known biosynthetic templates [142]. The exciting possibilities of this third approach have been demonstrated by the analysis of marine and freshwater ecosystems [143]. For example, sequencing 1.6 Gbp from the metagenome of the Sargasso Sea has revealed over 1.2 million new genes and 148 new microbial species [144]. The magnitude of sequence data released just in this study – totalling approximately 10 times the number of peptides currently archived in the UniProt database – dramatically illustrates the vast metabolic potential that metagenomics has to offer [145].

It is unlikely, even with detailed bioinformatic analysis that an entire metagenome could be handled in a manner similar to the bioinformatics process developed for annotating a solitary microbial genome. This is because the informatics challenges are too profound, for example, identification of genes and the *in silico* assembly of biosynthetic pathways must take into account the presence of multiple genomes at varying levels of coverage and degrees of relatedness. However, developing new algorithms to interrogate metagenomic sequences should allow construction of a single database that can theoretically contain all of the predicted open reading-frames (ORFs), their putative annotation, and biological role categories from a particular genomic source at a particular point in time. By examining the role assignments, it then becomes possible to identify all the biochemical pathways, and characterize the metabolic capacity of the entire community that may be present. Reconstructing as near-complete genome sequences as possible is especially important for gene expression analysis, with a view towards heterologous expression in a suitable host for sustainable compound supply. Integrative systems biology approaches blending *"-omics"* and bioinformatic analyses are likely to play a key role for this endeavour in the future [146].

Genome assembly from mixed sequences is readily achievable when a fully sequenced genome from a closely related source is available using alignment or comparative assembly programs [147,148]. However, as exemplified from the Sargasso Sea study, an appropriate template genome is usually not available and a scaffold of contiguous sequence separated by gaps of known size must be constructed by overlapping consensus assembly. This 'binning' step is heavily biased in that highly conserved sequences from phylogenetically distant DNA will cause false overlaps, whereas polymorphism within related DNA will introduce an artificial gap in the contiguous sequence [149]. Not surprisingly, community assembly is computationally very intensive, requiring manual post-processing of the 'binned' scaffolds to correct assembly errors [150]. Metagenomics for pharmaceutical lead discovery does not require the scaffolds to then be sorted into phylogenetic groups *per se*, but does require some degree of signature searching capacity to identify clones of interest. Novel gene discovery is the ultimate goal of this culture independent research.

We have been using hidden Markov models (HMM) [151] and family profiling programs such as HMMER database-building search software (http://hmmer.wustl.edu) [152] to predict catalytic functions of PKS and NRPS domains from gene sequences [38]. Preliminary work [153] has indicated that it is possible to construct HMMs to distinguish domains that incorporate common and more unusual biosynthetic precursors. The most difficult task has been to build HMMs that correctly predict the stereochemistry of the PKS and NRPS products, highlighting the limitations of existing bioinformatics techniques to assign absolute protein function based on expressed gene sequences – *Anfinsen's dogma* [154,155] (http://nobelprize.org/chemistry/laureates/1972/press.html). There is a need for new approaches to the computational problems associated with applying metagenomics for universal gene cloning and expression. These computational issues have been addressed in a recent overview [156].

## SCREENING METAGENOMIC LIBRARIES

Once a metagenomic library has been cloned, the task of screening the library commences. The first assumption is that a target metabolite of a marine invertebrate – microbial assemblage will be biosynthesized as a complete product from a cloned DNA fragment that encodes the entire biosynthetic pathway. This may not be the case, however, adding complexity to the deconvolution problem. For example, transacting elements may be encoded across separate clones necessitating completion of *in vitro* semi-synthesis from combined cell-free extracts. Many factors involved in recombinant selection have been considered [93,157]. It is recognised that the productive clones will be rare in any metagenomic library and that robust strategies must be developed to uncover the appropriate biosynthetic

clusters. Functional assays can be divided into types that select for an expressed trait: (a) gene assays that employ DNA primers to probe for the expected native biosynthetic pathways and (b) chemical assays to detect biosynthetic gene products based mainly on spectroscopic or spectrometric techniques often coupled to chemometric data analysis.

Initial success in this endeavour employed functional screens that distinguished a selectable phenotype, such as colour used for the detection of violacein and deoxyviolacein [124], or antibiosis for the detection of long-chain N-acyl amino acids from heterologously expressed soil DNA [158]. In the latter case, a simple *Bacillus subtilis* overlay assay was used to identify antibiotic producers through a zone of inhibition from which a library of 700,000 initial transformants yielded 65 antibiotic producing clones. Using an alternative approach, clones with antibiotic resistant genes were readily identified in a large soil metagenomic library by plating pools of clones onto antibiotic containing growth media, which gave 24 antibiotic resistant clones from an original library containing 5.4 Gbp of DNA [159]. More complex, functional assays have been developed to allow a very large throughput mainly to serve combinatorial chemistry programmes producing vast numbers of compounds. For example, the approach used by Huang and Schreiber [160] for selecting small molecule inhibitors of protein-protein interactions from a large combinatorial library was achieved using modified yeast cells in nanodroplets that could be applied to identify active clones in a metagenomic library. Other phenospecific assays such as those detecting metal chelation [161] can lack sensitivity and may not identify weakly productive clones. Additionally, conditions for a functional assay may not be compatible for growth by the surrogate host (the product itself could be toxic to the host), or procedures that require an intermediate step (i.e., solvent extraction) can limit the utility of the assay. Functional screens may not pinpoint the exact clone producing a particular metabolite of interest, yet a qualitative assay that is rapid to select for a phenotypic trait serves to reduce the number of clones that need to be examined by more rigorous methods.

Molecular techniques such as contig building, end sequence analysis, hybridisation and clone pooling assume the greatest importance after functional screening assays [162]. The successful use of hybridisation assays relies on several factors, such as the conserved nature of the biosynthetic genes, and the absence of similar biosynthetic sequences in the host genome. In the ideal situation, the genome of the metabolite-producing organism of interest has been fully sequenced. *A priori* knowledge of the *Prochloron didemni* genome was used by Schmidt *et al.* [27] to design unique primers to clone the patellamide biosynthetic pathway into *E. coli*. Such a strategy is costly and cannot be regarded as universal since it has limited use in the deconvolution of metagenomic clone libraries. PKS and NRPS gene clusters contain highly conserved sequences, but divergence may render degenerate primers based on terrestrial systems unsuitable to detect corresponding genes in marine organisms [93]. An implicit assumption is that a known type of biosynthetic pathway is being investigated. This is not always the case as was observed for the patellamide biosynthetic gene cluster where lack of hybridisation with two sets of degenerate primers for NRPS pathways necessitated a compound screening approach [105]. Automated instrumentation can now routinely examine large numbers of samples to enable compound-specific deconvolution of metagenomic clone libraries. As mentioned previously, preselection of clones may be executed based upon phenotypic or genotypic assays before compound screening is necessary, although in some cases direct chemical analysis may be the only possible option. In the case of cloning the patellamides from *Prochloron sp.*, deconvolution of a 1344-clone recombinant library was carried out by the culture of pooled clones, followed by solid phase extraction, liquid-liquid extraction and analyte concentration before analysis by high performance liquid chromatography with mass spectrometric detection (HPLC-MS) [105]. An HPLC-MS approach was also used to deconvolute a library of 1020 clones in a soil DNA library to obtain the metabolite terragines A-D, but no experimental details are provided [123].

Other methods needing minimal sample preparation that allow high throughput (<1 min/sample) are direct infusion electrospray ionisation mass spectrometric (ESI-MS) analysis of crude fermentation extracts and matrix-assisted laser desorption/ionization mass spectrometric (MALDI-MS) analysis of lyophilised cells embedded in a suitable matrix [163]. Both methods can be highly automated with use of liquid handling systems for high throughput, deconvolution of metagenomic DNA libraries. Analysis of ESI-MS and MALDI-MS data can involve a search for a single ion, or the data can be subjected to chemometric methods for metabolomic analysis [164]. This methodology, often using MS or nuclear magnetic resonance (NMR) spectroscopy data obtained on 'raw' sample extracts, allows a comparative assessment of chemotypic differences between clones in a recombinant DNA library. Current liquid handling systems and probe designs allow the acquisition of more than one proton ($^1$H) NMR spectrum per minute, potentially making NMR an efficient method to pre-screen metagenomic clone libraries for secondary metabolites.

## CONCLUSION

While the technology for cloning the genes responsible for the biosynthetic assembly of marine metabolites for heterologous expression in a surrogate host is unequivocally still in its infancy, the way is clear for significant advancements. Adopting a molecular approach will allow genetic manipulations of key assembly pathways for the construction of combinatorial biosynthetic libraries [106,165,166] that portends the rational design of hybrid metabolites for novel drug discovery. Further work in this field may additionally enable the biosynthetic expression of all natural products encoded in the metagenome of a marine invertebrate and its complex microbial consortia, not just those that are biosynthetically expressed under a particular set of environmental conditions and metabolic challenges (stress, infection, etc). Adapting a molecular approach will allow probing the unique nature of invertebrate-microbial associations in the bioprocess engineering and ecological functioning of marine secondary metabolites [167] that entail partner-exclusive, interactive, or shared metabolic pathways of the holobiont. Overall, this strategy will only succeed

through the concomitant development of appropriate bioinformatic tools and high-throughput methodologies to screen the clone libraries that are generated in all the proposed approaches to select for the desired production of a marine bioactive metabolite.

## ACKNOWLEDGEMENTS

We extend our thanks to Dr. David Newman of the National Cancer Institute for his critical reading of this manuscript. Financial support for this work has come from the Australian Institute of Marine Science; The Ministry of Science, Education and Sport, Republic of Croatia; and from the United Kingdom by the College of Physical Sciences University of Aberdeen, The University of London School of Pharmacy, the Royal Society of Edinburgh, the Royal Society of London; and the Carnegie Trust for the Universities of Scotland, The Welcome Trust and The Leverhulme Trust.

## ABBREVIATIONS

| | | |
|---|---|---|
| TEM | = | Transmission electron micrograph |
| NRPS | = | Nonribosomal peptide synthetase |
| PKS | = | Polyketide synthase |
| $PLA_2$ | = | Phospholipase $A_2$ |
| TPS | = | Terpene synthase |
| BAC | = | Bacterial artificial chromosome |
| ORFs | = | Open reading-frames |
| HMM | = | Hidden Markov models |
| HPLC-MS | = | High performance liquid chromatography and mass spectrometric detection |
| ESI-MS | = | Electrospray ionisation mass spectroscopy |
| MALDI-MS | = | Matrix-assisted laser desorption/ionization mass spectrometry |
| NMR | = | Nuclear magnetic resonance [spectroscopy] |
| $^1H$ | = | Proton |

## REFERENCES

[1] Rouhi, A.M. *Chem. Eng. News,* **1995**, 42.
[2] Battershill, C.N.; Jaspars, M.; Long, P.F. *Biologist* **2005**, *52*, 107.
[3] Firn, R.D.; Jones, C.G. *Nat. Prod. Rep.* **2002**, *30*, 382.
[4] Newman, D.J.; Cragg, G.M. *J. Nat. Prod.* **2004**, *67*, 1216.
[5] Newman, D.J.; Cragg, G.M. *Curr. Med. Chem.* **2004**, *11*, 1693.
[6] Munro, M.H.; Blunt, J.W.; Dumdei, E.J.; Hickford, S.J.; Lill, R.E.; Li, S.; Battershill, C.N.; Duckworth, A.R. *J. Biotechnol.* **1999**, *70*, 15.
[7] Friedrich, A.B.; Merkert, H.; Fendert, T.; Hacker, J.; Proksch.; Hentschel, U. *Mar. Biol.* **1999***, 134*, 461.
[8] Bergquist, P. R. Hutchinson and Company, London, **1978** pp. 268.
[9] Webster, N.S.; Hill, R.T. *Mar. Biol.* **2001**, *138*, 843.
[10] Webster, N.S.; Wilson, K.J.; Blackall, L.L.; Hill, R.T.P. *Appl. Environ. Microbiol.* **2001**, *67*, 434.
[11] Hentschel, U.; Hopke, J.; Horn, M.; Friedrich, A.B.; Wagner, M.; Hacker, J.; Moore, B.S. *Appl. Environ. Microbiol.* **2002**, *68*, 4431.
[12] Taylor, M.W.; Schupp, P.J.; Dahilöf, I.; Kjelleberg, S.; Steinberg, P.D. *Environ. Microbiol.* **2004**, *6*, 121.
[13] Webster, N.S.; Negri, A.P.; Munro, M.M.; Battershill, C.N. *Environ. Microbiol.* **2004**, *6*, 288.
[14] Erpenbeck, D.; Breeuwer, J.A.J.; van der Velde, H.C.; van Soest, R.W.M. *Mar. Biol.* **2002**, *141*, 377.
[15] Thacker, R.W.; Starnes, S. *Mar. Biol.* **2003**, *142*, 643.
[16] Gunasekera, S.P.; Gunasekera, M.; Longley, R.E. *J. Org. Chem.* **1990**, *55*, 4912.
[17] Proksch, P.; Edrada, A.; Ebel, R. *Appl. Microbial. Biotechnol.* **2002**, *59*, 125.
[18] Haygood, M.G.; Davidson, S.K. *Env Microbiol* **1997**, *63*, 4612.
[19] Davidson, S.K.; Allen, S.W.; Lim, G.E.; Anderson, C.M.; Haygood, M.G. *Appl. Environ. Microbiol.* **2001**, *67*, 4531.
[20] Hildebrand, M.; Waggoner, L.E.; Lim, G.E.; Sharp, K.H.; Ridley, C.P.; Haygood, M.G. *Nat. Prod. Rep.* **2004**, *21*, 122.
[21] Piel, J. *Nat. Prod. Rep.* **2004**, *21*, 519.
[22] Li, A.; Piel, J. *Chem. Biol.* **2002**, *9*, 1017.
[23] Degnan, B.M.; Hawkins, C.J.; Lavin, M.F.; McCaffrey, E.J.; Parry, D.L.; van den Brenk, A.L.; Watters, G.J. *J. Med. Chem.* **1989**, *32*, 1349.
[24] Ishida, T.; In, Y.; Doi, M.; Inoue, M.; Hamada, Y.; Shioiri, T. *Biopolymers* **1992**, *32*, 131.
[25] Salomon, C.E.; Faulkner, D.J. *J. Nat. Prod.* **2002**, *65*, 689.
[26] Schmidt, E.W.; Sudek, S.; Haygood, M.G. *J. Nat. Prod.* **2004**, *67*, 1341.
[27] Schmidt, E.W.; Nelson, J.T.; Rasko, D.A.; Sudek, S.; Eisen, J.A.; Haygood, M.G.; Ravel, J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7315.
[28] van den Brenk, A.L.; Tyndall. J.D.; Cusack, R.M.; Jones, A.; Fairlie, D.P.; Gahan, L.R.; Hanson, G.R. *J. Inorg. Biochem.* **2004**, *98*, 1857.
[29] Williams, A.D.; Jacobs, R.S. *Cancer Lett.* **1993**, *71*, 97.
[30] Cane, D.E.; Walsh, C.J. *Chem. Biol.* **1999**, *6*, 319.
[31] Mootz, H.D.; Schazer, D.; Marahiel, H.A. *Chem. Bio. Chem.* **2002**, *3*, 490.
[32] Piel, J.; Hui, D.; Fusetani, N.; Matsunaga, S. *Environ. Microbiol.* **2004**, *6*, 921.
[33] Schirmer, A.; Gadkari, R.; Reeves, C.D.; Ibrahim, F.; DeLong, E.F.; Hutchinson, C.R. *Appl. Environ. Microbiol.* **2005**, *71*, 4840.
[34] Staunton, J.; Weissman, K.J. *Nat. Prod. Rep.* **2001**, *18*, 380.
[35] McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Koshla, C. *Science* **1993**, *262*, 1546.
[36] Hranueli, D.; Peric, N.; Borovička , B.; Bogdan, S.; Cullum, J.; Waterman, P.G.; Hunter, I.A. *Food Technol. Biotechnol.* **2001**, *39*, 203.
[37] Donadio, S.; Sosio, M. *Comb. Chem. High Throughput Screen* **2003**, *6*, 489.
[38] Hranueli, D.; Cullum, J.; Basrak, B.; Goldstein, P.; Long P.F. *Curr. Med. Chem.* **2005**, *12*, 1697.
[39] Perry, N.B.; Blunt, J.W.; Munro, M.H.G.; Pannell, L.K. *J. Am. Chem. Soc.* **1988**, *110*, 4850.
[40] West, L.M.; Northcote, P.T.; Hood, K.A.; Miller, J.H.; Page, M.J. *J. Nat. Prod.* **2000**, *63*, 707.
[41] Burres, N.S.; Clement, J.J. *Cancer Res.* **1989**, *49*, 2935.
[42] Hood, K.A.; West, L.M.; Northcote, P.T.; Berridge, M.V.; Miller, J.H. *Apoptosis* **2001**, *6*, 155.
[43] Romo, D.; Rzasa, R.M.; Shea, H.A.; Park, K.; Langehan, J.M.; Sun, L.; Akheizer, A.; Liu, J.O. *J. Am. Chem. Soc.* **1998**, *120*, 12237.
[44] Northcote, P.T.; Blunt, J.W.; Munro, M.H.G. *Tetrahedron Lett.* **1991**, *32*, 6411.
[45] West, L.M.; Northcote, P.T.; Battershill, C.N. *J. Org. Chem.* **2000**, *65*, 445.
[46] Hood, K.A.; West, L.M.; Rouwe, B.; Northcote, P.Y.; Berridge, M.V.; Wakefield, S.J.; Willer, J.H. *Cancer Res.* **2002**, *62*, 3356.
[47] Hay, M.E. *J. Exp. Mar. Biol. Ecol.* **1996**, *200*, 103.
[48] Page, M.; West, L.; Northcote, P.; Battershill, C.; Kelly, M. *J. Chem. Ecol.* **2005**, *31*, 1161.
[49] Webb, V.L.; Maas, E.W. *FEMS Microbiol. Lett.* **2002**, *207*, 43.
[50] Trapp, S.C.; Croteau, R.B. *Genetics* **2001**, *158*, 811.
[51] Masse, G.; Belt, S.T.; Rowland, S.J.; Rohmer, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4413.
[52] Martin, V.J.; Pitera, D.J.; Withers, S.T.; Newman, J.D.; Keasling, J.D. *Nat. Biotechnol.* **2003**, *21*, 796.
[53] de Silva, E.D.; Scheuer, P.J. *Tetrahedron Lett.* **1990**, *21*, 1611.
[54] de Silva, E.D.; Scheuer, P.J. *Tetrahedron Lett.* **1981**, *22*, 3147.
[55] Jacobs, R.S.; Culver, P.; Langdon, R.; O'Brien, T.; White, S. *Tetrahedron* **1985**, *41*, 981.

[56] Glaser, K.B.; Jacobs, R.S. *Biochem. Pharmacol.* **1986**, *35*, 449.
[57] Glaser, K.B.; Jacobs, R.S. *Biochem. Pharmacol.* **1987**, *36*, 2079.
[58] Glaser, K.B.; De Carvalho, S.M.; Jacobs, R.S.; Kernan, M.R.; Faulkner, D.J. *Mol. Pharmacol.* **1989**, *36*, 782.
[59] Jacobson, P.B.; Marshall, L.A.; Sung, A.; Jacobs, R.S. *Biochem. Pharmacol.* **1990**, *39*,1557.
[60] Wheeler, L.; Sachs, G.; De Vries, G.; Doodrum, D.; Woldemussie, E.; Muallem, S. *J. Biol. Chem.* **1987**, *262*, 6531.
[61] Kernan, M.R.; Faulkner, D.J.; Jacobs, R. S. *J. Org. Chem.* **1987**, *52*, 3081.
[62] Albizati, K.F.; Holman, T.; Faulkner D.J.; Glaser K.B.; Jacobs, R.S. *Experientia* **1987**, *43*, 949.
[63] Kernan, M.R.; Faulkner, D.J.; Parkanyi, L.; Clardy, J.; de Carvalho, M.S.; Jacobs, R.S. *Experientia* **1989**, *45*, 388.
[64] Butler, M.S.; Capon R.J. *Aust. J. Chem.* **1992**, *45*, 705.
[65] Potts, B.C.M.; Faulkner, D.J.; De Carvalho, M.S.; Jacobs, R.S. *J. Am. Chem. Soc.* **1992**, *114*, 5093.
[66] de Rosa, M.; Giordana, S.; Scettri, A.; Sodano, G.; Soriente, A.; Garcia-Pastor, P.; Alcarez, M.J.; Paya, M. *J. Med. Chem.* **1998**, *41*, 3232.
[67] Tsuda, M.; Endo, T.; Mikami, Y.; Fromont, J.; Kobayashi, J. *J. Nat. Prod.* **2002**, 65, 1507.
[68] Namikoshi, M.; Suzuki, S.; Meguro, S.; Nagal, H.; Koike, Y.; Kitazawa, A.; Kobayashi, H.; Oda, T.; Yamada, J. *Fisheries Science* **2004**, *70*, 152.
[69] Faulkner, D.J. *Antonie van Leeuwenhoek* **2000**, *77*, 135.
[70] Katsumura, S.; Fujiwara, S.; Isoe, S. *Tetrahedron Lett.* **1985**, *26*, 5827.
[71] Katsumura, S.; Fujiwara, S.; Isoe, S. *Tetrahedron Lett.* **1988**, *29*, 1173.
[72] Garst, M.E.; Tallman, E.A.; Bonfiglio, J. N.; Harcourt, D.; Ljungwe, E.B.; Tran, A. *Tetrahedron Lett.* **1986**, *27*, 4533.
[73] Bury, P.; Hareau-Vittini, G.; Kocienski, P.; Dhanak, D. *Tetrahedron* **1994**, *50*, 8793.
[74] Pommier, A.; Kocienski, P.J. *Chem. Commun.* **1997**, *12*, 1139.
[75] Soriente, A.; de Rosa, M.; Scettri, A.; Sodano, G.; Terencio, M.C.; Paya, M.; Alcaraz, M.J. *Curr. Med. Chem.* **1999**, *6*, 415.
[76] Pommier, A.; Stepanenko, V.; Jarowicki, K.; Kocienski, P.J. *J. Org. Chem.* **2003**, *68*, 4008.
[77] Kohl, A.C.; Ata, A.; Kerr, R.C. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 495.
[78] Kohl, A.C.; Kerr, R.G. *Arch. Biochem. Biophys.* **2004**, *424*, 97.
[79] Ferns, T.A.; Kerr, R.G. *J. Org. Chem.* **2005**, *70*, 6125.
[80] Mydlarz, L.D.; Jacobs, R.S.; Boehnlein, J.; Ker, R.G. *Chem. Biol.* **2003**, *10*, 1051.
[81] Proksch, P.; Edrada, A.; Ebel, R. *Mar. Drugs* **2003**, *1*, 5.
[82] Francavilla, C.; Chen, W.; Kinder, F.R. Jr. *Org. Lett.* **2003**, *5*, 1233.
[83] Mickel, S.J. *Curr. Opin. Drug Discov. Dev.* **2004**, 7, 869.
[84] Pettit, G.R.; Kamano, Y.; Herald, C.L.; Tuinman, A.A.; Boettner, F.E.; Kizu, H.; Schmidt, J.M.; Baczynskyj, L.; Tomer, K.B.; Bontems, R. *J. Am. Chem. Soc.* **1987**, *109*, 6883.
[85] Martinez, E.J.; Cory, E.J. *Org. Lett.* **2000**, *2*, 993.
[86] Cuevas, C.; Perez, M.; Martin, M.J.; Chicharro, J.L.; Fernández-Rivas, A.; Flores, M.; Francesch, A.; Gallego, P.; Zarzuelo, M.; de La Calle, F.; Garcia, J.; Polanco, C.; Rodrigez, I., Manzanares, I. *Org. Lett.* **2000**, *10*, 2545.
[87] Battershill, C.N.; Page, M.J.; Duckworth, A.R.; Miller, K.A.; Bergquist, P.R.; Blunt, J.W.; Munro, M.H.G.; Northcote, P.T.; Newman, D.J.; Pomponi, S.A. *Abstracts, 5th International Sponge Symposium* Brisbane, **1998**, pp. 16.
[88] Hart, J.B.; Lill, R.E.; Hickford, S.J.H.; Blunt, J.W.; Munro, M.H.G. *Drugs from the Sea*; Fusetani, Ed.; Karger: Basel, **2000**, pp. 134.
[89] Mendola, D. Aquaculture production of bryostatin 1 and ecteinascidin 743. *Drugs from the Sea*; Fusetani, Ed.; Karger: Basel, **2000**; pp. 120.
[90] Mendola, D. *Biomol. Eng.* **2003**, *20*, 441.
[91] Tagliatela-Scafati, O.; Deo-Jangra, U.; Cambell, M.; Roberge, M.; Andersen, R.J. *Org. Lett.* **2002**, *4*, 4085.
[92] Piel, J.; Hui, D.; Wen, G.; Butzke, D.; Platzer, M.; Fusetani, N.; Matsunaga, S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16222.
[93] Fortman, J.L.; Sherman, D.H. *Chem. Bio. Chem.* **2005**, *6*, 1.
[94] Ueno, T.; Takahashi, H.; Oda, M.; Mizunuma, M.; Yokoyama, A.; Goto, Y.; Mizushima, Y.; Sakaguchi, K.; Hayashi, H. *Biochem.* **2000**, *39*, 5995.
[95] Kasanah, N.; Hamann, M.T. *Curr. Opin. Investig. Drugs* **2004**, *5*, 827.
[96] Hill, R.T.; Hamann, M.T.; Peraud, O.; Kasanah, N. Patent WO 2004/013297, University of Maryland Biotechnology and University of Mississippi, 12 February **2004.**
[97] Schroder, H.C.; Brummer, F.; Fattorusso, F.; Aiello, A.; Menna, M.; de Rosa, S.; Batel, R.; Muller, W.E. *Prog. Mol. Subcell Biol.* **2003**, *37*, 163.
[98] Muller, W.E.; Grebenjuk, V.A.; Le Pennec, G.; Schroder, H.C.; Brummer, F.; Hentschel, U.; Muller, I. M.; Breter, H.J. *Mar. Biotech.* (NY) **2004**, *6*, 105.
[99] Sipkema, D.; Osinga, R.; Schatton, W.; Mendola, D.; Tramper, J.; Wijffels, R.H. *Biotechnol. Bioeng.* **2005**, *90*, 201.
[100] Piel, J.; Hertweek, C.; Shipley, P.R.; Hunt, D.M.; Newman, M.S.; Moore, B.S. *Chem. Biol.* **2000**, *7*, 943.
[101] Chang, Z.; Flatt, P.; Gerwick, W.H.; Nguyen, V.A.; Willis, C.L.; Sherman, D.H. *Gene* **2002**, *296*, 235.
[102] Edwards, D.J.; Gerwick, W.H. *J. Am. Chem. Soc.* **2004**, *126*, 11432.
[103] Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien B. *Science* **2000**, *267*, 640.
[104] Boddy, C.N.; Hotta, K.; Tse, M.L.; Watts, R.E.; Khosla, C. *J. Am. Chem. Soc.* **2004**, *126*, 7436.
[105] Long, P.F.; Dunlap, W.C.; Battershill, C.N.; Jaspars, M. *Chem. Bio. Chem.* **2005**, *6*, Jun 29; Epub ahead of print.
[106] Walsh, C.T. *Science* **2004**, *303*, 1805.
[107] Weber, T.; Welzel, K.; Pelzer, S.; Vente, A.; Wohlleben, W. *J. Biotechnol.* **2003**, *106*, 221.
[108] Goodfellow, M.; Haynes, J.A. *Biological, Biochemical, and Biomedical Aspects of Actinomycetes* Oritz-Oritiz, Bojalil, Yakoleff, Eds; Academic Press: New York, **1984**; pp.453.
[109] Helmke, E.; Weyland, H. *Int. J. Syst. Bacteriol.* **1984**, *34*, 127.
[110] Jensen, P.R.; Dwight, R.; Fenical, W. Distribution of actinomycetes in near-shore tropical marine sediments. *Appl. Environ. Microbiol.* **1991**, *57*, 1102.
[111] Takizawa, M. Colwell, R.A.; Hill, R. Isolation and diversity of actinomycetes in the Chesapeak Bay. *Appl. Environ. Microbiol.* **1993**, *59*, 997.
[112] Moran, M.A.; Rutherford, L.T.; Hodson, R.E. *Appl. Environ. Microbiol.* **1995**, *61*, 3695.
[113] Colquhoun, J.A.; Mexson, J.; Goodfellow, M.; Ward, A.C.; Horikoshi, K.; Bull, A.T. *Antonie Van Leeuwenhoek* **1998**, *74*, 27.
[114] Bull, A.T.; Stach, J.E.; Ward, A.C.; Goodfellow, M. *Antonie Van Leeuwenhoek* **2005**, *87*, 65.
[115] Mincer, T.J.; Jensen, P.R.; Kaufmann, C.A.; Fenical, W. *Appl. Environ. Microbiol.* **2002**, *61*, 5005.
[116] Jensen, P.R.; Gontang, E.; Mafnas, C.; Mincer, T.J.; Fenical, W. *Environ. Microbiol.* **2005**, *7*, 1039.
[117] Kim, T.K.; Garson, M.J.; Fuerst, J.A. *Environ. Microbiol.* **2005**, *7*, 509.
[118] Moore, B.S.; Kalaitzis, J.A.; Xiang, L. *Antonie Van Leeuwenhoek* **2005**, *87*, 49.
[119] Donadio, S.; Monciardini, P.; Alduina, R.; Mazza, P.; Chiocchini, C.; Cavalette, L.; Sosio, M.; Puglia, A.M. *J. Biotechnol.* **2002**, *99*, 187.
[120] Handelsman, J. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669.
[121] Handelsman, J.; Rondon, M.R.; Brady, S.F.; Clardy, J.; Goodman, R.M. *Chem. Biol.* **1998**, *5*, R245.
[122] Rondon, M.R.; August, P.R.; Bettermann, A.D.; Brady, S.F.; Grossman, T.H.; Liles, M.R.; Loiacono, K.A.; Lynch, B.A.; MacNeil, I.A.; Minor, C.; Tiong, C.L.; Gilman, M.; Osbourne, M.S.; Clardy, J.; Handelsman, J.; Goodman, R.M. *Appl. Environ. Microbiol.* **2000**, *66*, 2541.
[123] Wang, G-Y-S.; Graziani, E.; Waters, B.; Pan, W.; Li, X..; McDermott, J.; Meurer, G.; Saxena, G.; Andersen, R.J.; Davies, J. *Org. Lett.* **2000**, *2*, 2401.
[124] Brady, S.F.; Chao, C.; Handelsmann, J.; Clardy, J. *Org. Lett.* **2001**, *3*, 1981.
[125] Gillespie, D.E.; Brady, S.F.; Bettermann, A.D.; Cianciotto, N.P.; Liles, M.R.; Rondon, M.R.; Clardy, J.; Goodman, R.M.; Handelsman, J. *Appl. Environ. Microbiol.* **2002**, *68*, 4301.
[126] Courtois, S.; Cappellano, C.M.; Ball, M.; Francou, F.-X.; Normand, P.; Helynck, G.; Martinez, A.; Kolvek, S.J.; Hopke, J.; Osburne, M.; August, P.R.; Nalin, R.; Guérineau, M.; Jeannin, P.; Simonet, P.; Pernodet, J.-L. *Appl. Environ. Microbiol.* **2003**, *69*, 49.
[127] Schloss, P.D.; Handelsman, J. *Genome Biol.* **2005**, *6*, 229.

[128] Béjà, O.; Koonin, E.V.; Aravind, L.; Taylor, L.T.; Seitz, H.; Stein J.L.; Bensen, D.C.; Feldman, R.A.; Swanson, R.V.; Delong, E.F. *Appl. Environ. Microbiol.* **2000**, *68*, 335.
[129] Béjà, O.; Susuki, M.T.; Koonin, E.V.; Aravind, L.; Hadd, A.; Nguyen, L.P.; Villacorta, R.; Amjada, M.; Garrigues, C.; Jovanovich, S.B.; Feldman, R.A.; DeLong, E.F. *Environ. Microbiol.* **2002**, *2*, 516.
[130] Sosio, M.; Giusino, F.; Cappellano, C.; Bossi, E.; Puglia, A.M.; Donadio, S. *Nat. Biotechnol.* **2000***, 18*, 343.
[131] Lopanik, N.; Lindquist, N.; Targett, N. *Oecologia* **2004**, *139*, 131.
[132] Gabor, E.M.; Alkema, W.B.L.; Jenssen, D.B. *Environ. Microbiol.* **2004**, *6*, 879.
[133] Daly, R.; Haern, M.T. *J. Mol. Recognit.* **2005**, *18*, 119.
[134] Baum, D.A.; Testa, S.M. *RNA* **2005**, *11*, 897.
[135] Pfeifer, B.A.; Admiraal, S.J.; Gramajo, H.; Cane, D.E.; Khosla, *Science* **2001**, *291*, 1790.
[136] Sorensen, H.P.; Mortensen, K.K. *J. Biotechnol.* **2005***, 102*, 113.
[137] Martinez, A.; Kolvek, S.J.; Yip, C.L.T.; Hopke, J.; Brown, K.A.; MacNeil, I.A.; Osburne, *Appl. Environ. Microbiol.* **2004**, *70*, 2452.
[138] Streitand, W.R.; Schmitz, R.A. *Curr. Opin. Microbiol.* **2004**, *7***,** 492.
[139] Uchiyama, T.; Abe, T.; Ikemura, T.; Watanabe, K. *Nat. Biotechnol.* **2005**, *23*, 88.
[140] Stein, J.L.; Marsh, T.L.; Wu, K.Y.; Shizuya, H.; DeLong, E.F. *J. Bacteriol.* **1996**, *178*, 591.
[141] Béjà, O.; Aravind, L.; Koonin, E.V.; Suzuki, M.T.; Hadd. A.; Nguyen, L.P.; Jovanovich, S.B.;Gates, C.M.; Feldman R.A.; Spudich, E.N.; DeLong, E.F. *Science* **2000**, *289*, 1902.
[142] Riesenfeld, C.S.; Schloss, P.; Handelsman, J. *Annu. Rev. Genet.* **2004**, *38*, 525.
[143] Tringe, S.; von Mering, C.; Kobayashi, A.; Salamov, A.; Chen K.; Chang, H.W.; Podar, M.; Short, J.M.; Mathur E.J.; Detter, J.C.; Bork, P.; Hugenholtz, P.; Ruben, E.M. *Science* **2005**, *308*, 554.
[144] Venter, J.C.; Remington, K.; Heidelberg, J.F.; Halpern, A.L.; Rusch, D.; Eisen, J.A.; Wu, D.; Paulsen, I.; Nelson, K.E.; Fouts, D.E.; Levy, S.; Knap, A.H.; Lomas, M.W.; White, O.; Peterson, J.; Hoffman, J.: Parsons, R.; Baden-Tilson, H.; Pfannkoch, C.; Rogers, Y.H.; Smith, H.O. *Science* **2004**, *304*, 66.
[145] Falkowski, P.G.; de Vargas, C. *Science* **204**, *304*, 58.
[146] DeLong, E.F. *Environ. Microbiol.* **2004**, *6*, 875.
[147] Pop, M.; Philippy, A.; Delcher, A.L.; Salzberg, S.L. *Brief Bioinform.* **2004**, *5*, 237.
[148] Bray, N.; Pachter, L. MAVID: *Genome Res.* **2004**, *14*, 693.
[149] Breitbart, M.; Salamon, P.; Andresen, B.; Mahaffy. J.; Segal, A.M.; Azam, F.; Rohwer, F. *Proc. Natl. Acad. Sci. USA* **2002,** *99*, 14250.
[150] DeLong, E.F. *Nat. Rev. Microbiol.* **2005**, *3*, 459.
[151] Durbin, R.; Eddy, S; Krogh A.; Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press: Cambridge, **2001**.
[152] Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755.
[153] Goldstein, P.; Basrak, B.; Žučko , J.; Starčević, A.; Hranueli, D.; Long, P.F.; & Cullum J. *Programme and Abstracts of the 4th Scientific Conference on Biotechnology: Biotechnology and Immuno-Modulatory Drugs*, Zagreb, **2005**, pp. 32.
[154] Anfinsen, C.B. *Biochem. J.* **1972**, 128, 737.
[155] Anfinsen, C.B. *Science* **1972**, *96*, 223.
[156] Chen, K.; Pachter, L. *PLOS Comput. Biol.* **2005**, *2*, e24.
[157] Schloss, P.D.; Handelsman, J. *Curr. Opin. Biotechnol.* **2003**, *14*, 303.
[158] Brady, S.F.; Clardy, J. *Org. Lett.* **2005**, *7*, 3613.
[159] Riesenfeld, C.S.; Goodman, R.M.; Handelsman, J. *Environ. Microbiol.* **2004**, *6*, 981.
[160] Huang, J., Schreiber, S. A. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 13396.
[161] Schwyn, B.; Neilands, S.B. *Anal. Biochem*. **1987**, *160*, 47.
[162] Rondon, M.R.; Raffel, S.J.; Goodman, R.M.; Handelsman, J. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6451.
[163] Erhard, M.H.; von Dohren, H.; Jungblut, P. *Nat. Biotech.* **1997**, *15*, 906.
[164] Nicholson, J.; Connelly, J.; Lindon, J.C.; Holmes, E. *Nat. Rev. Drug Disc.* **2002**, *1*, 153.
[165] Sosio, M.; Bossi, E.; Donadio, S. *Nucleic Acids Res.* **2001**, *37*, e37.
[166] Rix, U.; Fisher, C.; Remsing, L.L.; Rohr, U. *Nat. Prod. Rep.* **2002**, *19*, 542.
[167] Osinga, R.; Armstrong, E.; Burgess, J.G.; Hoffman, F.; Reitner, J.; Schumann-Kindel, G. *Hydrobiologia* **2001**, *461*, 55.

Received: September 14, 2005 Revised: November 29, 2005 Accepted: November 30, 2005