Drug discovery from uncultivable microorganisms

Brajesh K. Singh[1,2,] and Catriona A. Macdonald[3]

[1] The Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK
[2] Centre for Plants and the Environment, University of Western Sydney, Penrith South, DCNSW 1797, Australia
[3] Rothamsted Research, Rothamsted, Harpenden, AL5 2JQ, UK

Environmental microbes are a major source of drug discovery, and several microbial products (antibiotics, anti-tumour products, immunosuppressants and others) are used routinely for human therapies. Most of these products were obtained from cultivable (<1%) environmental microbes, and this means that the vast majority of microbes were not targeted for drug discovery. With the advent of new and emerging technologies, we are poised to harvest novel drugs from the so-called 'uncultivable' microbes. In this article, we propose how a multidisciplinary approach combining different technologies can expedite and revolutionize drug discovery from uncultivable microbes and examine the current limitations of technologies and strategies to overcome such limitations that might further expand the promise of drugs from environmental microbes.

Introduction

Natural products are a major source of drug discovery, accounting for 60% of the total market. Despite the 30% drop in natural-product-based drugs in clinical studies between 2001 and 2008 [1], natural products remain the main source of drug supply: more than 40% of new drugs discovered since 1980 are derived from natural microbial sources [2]. In addition, 70% of antibiotics in current use are either natural products or derivatives of natural products [3]. In the same period, 74% of all new chemical entities for cancer, 70% of anti-migraine drugs and 66% of anti-hypertensive agents were derived from natural products [4]. Despite previous success, the recent trend is discouraging: most pharmaceutical companies are decreasing their screening efforts for natural products. Traditionally, the successful 'hit rate' for new drugs from screening natural products has been high. For example, 20 commercial drugs were obtained from screening just over 7000 known structures for polykatides – a hit rate of 0.3%, which is much higher than the <0.001% hit rate from high-throughput screening (HTS) of synthetic compound libraries [1]. Despite this, the pharmaceutical industry favours the HTS of massive libraries from pure synthetic compounds, mainly because they are screen friendly and allow for a short timeline in screening for a large number of molecules. HTS for natural products, however, suffers from several shortcomings: reliable access and supply, concerns regarding intellectual property arising under the Rio Convention on Biodiversity, and the loss of natural sources caused by habitat destruction (urbanization, intensive farming, flood, draught, and climate change) [1,3].

Microbes as a source of drugs

Microbes isolated from the natural environment are an amazing source of diverse drugs. Competition for survival and environmental pressures drive the evolution of defence, attack and signalling diversity, which, in turn, determine chemical and biological diversity and potential new drugs. Environmental microbes – mainly actinomycetes, bacilli and filamentous fungi – have an enormous capacity to produce secondary metabolites, which have been exploited for drug discovery. Natural products obtained from microbes have a long history. For example, *Aspergillus oryzae* has been used for making koji from rice for 5000 years, and *Penicillium roqueforti* has been employed to make cheese for 4000 years [4]. By 2002, microbes were the source for 22,500 bioactive compounds. Of these, 17% were obtained from unicellular bacteria (mainly *Pseudomonas* and *Bacillus*), 45% from filamentous bacteria

Corresponding author:. Singh, B.K. (b.singh@uws.edu.au)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.07.002

TABLE 1

A list of some recently discovered drugs derived from microbial products.

Compound	Source	Activity
Drugs approved for medical treatment		
Amrubicin hydrochloride	*Streptomyces peucitius*	Anti-cancer
Gemtuzumab ozogamycine	*Micromonospora echinospora*	Anti-cancer
Rosuvastatin calcium	*Penicillium citrinum*	Lipid-lowering
Pimecrolimus	*Streptomyces hygroscopicus*	Anti-atopic dermatitis
Everolimus	*Streptomyces hygroscopicus*	Immunosuppressant
Mycophenolate sodium	*Penicillium brevicompactum*	Immunosuppressant
Miglustat	*Streptomyces lavendulae*	Gaucher disease type I
Biapenem	*Streptomyces cattleya*	Antibacterial
Deptomycine	*Streptomyces raseosporus*	Antibacterial
Telithromycine	*Saccaropolyspora erythraea*	Antibacterial
Drugs under clinical trials		
Becatecarin	*Lechevalieria aerocolonigenus*	Anti-cancer
Eco-4601	*Micromonospora* sp.	Anti-cancer
NPI-0052	*Salinispora tropica*	Anti-cancer
Elsamirucini	*Streptomyces chartreusis*	Anti-cancer
Romidepsin	*Chromobacterium violaceum*	Anti-cancer
Spicamycine	*Streptomyces alanosinicus*	Anti-cancer
Vorinostat	*Streptomyces hygroscopicus*	Anti-cancer
Ramoplanin	*Actionoplanes* sp.	Anti-infective
Retapamulen	*Pleurotus mutiliz*	Anti-infective
Tiacumicin	*Dactylosporungium aurantiacum*	Anti-infective

(actinomycetes) and 38% from fungi [4]. The major areas of research in human health in relation to microbial drugs include diabetes, anti-inflammatory, haematology and cancer and respiratory, gastro-intestinal, infectious, central nervous system and cardiovascular diseases [3]. Several currently used antibiotics, immunosuppressants and anti-cancer drugs were isolated from environmental microbes [4,5] (Table 1). More than 100 natural products are currently undergoing clinical trials, and a similar number of products are at the preclinical stage [2,3,5]. According to various estimates, of the 25 top-selling drugs in 1997, 42% originated from natural products. Antibiotics constitute 67% of such sales and had a worldwide market value of $32 billion in 2001. Antibiotics from *Streptomyces* alone had a sales value of $25 billion [6].

Emerging microbial drugs

Microbes can also be used as the main source of emerging therapies, such as in the use of protein and medical bioremediation. The use of protein therapy has increased exponentially since the first use of human insulin 25 years ago. At present, more than 130 protein therapeutics are used worldwide and many more are at the developmental or clinical trial stage [7]. Besides obvious clinical benefits, the development of protein therapeutics is considered financially more attractive than chemical or natural products by some drug companies because the clinical development and Federal Drug Authority approval time are at least one year faster than those for small-molecule drugs developed between 1980 and 2002 [8] and because most of the proteins are novel in terms of structure and function, which enables hassle-free patent acquisition [7]. At present, the contribution of microbial proteins to human therapy is increasing and possesses the potential to revolutionize human treatment in several areas of medicine.

Medical bioremediation is an emerging field that involves the treatment of several metabolic conditions using microbial enzymes [9]. Because catabolic insufficiency is one of the main causes of aging, it has been proposed that the removal of specific metabolic products that accumulate intracellularly would remove symptoms of some diseases or aging. Medical bioremediation, which involves the use of exogenous enzymes to counteract the effect of missing or insufficient enzymes involved in metabolic reactions, has been suggested as one option for such treatment [10]. For example, 7-ketocholesterol which has cytotoxic properties, has been associated with Alzheimer's disease and atherosclerosis [11]. It has been shown that reducing the concentration of this molecule results in the slowing of atherosclerotic progression and, therefore, 7-ketocholesterol is considered an ideal target for medical bioremediation. In a recent study, several bacteria that can mineralize 7-ketocholesterol were isolated [10]. Similarly, degradation of amyloid beta and neurofibrillary tangles (two neuropathological hallmarks for Alzheimer's disease) by *Mycoplasma hyorhinis* and *Mycoplasma penetrans* were observed [12].

Likewise, the options for treating chemical poisoning are limited. For example, organophosphorus (OP) compounds – widely used pesticides – account for 38% of global pesticide use, are highly toxic to mammals and have been reported to be responsible for approximately three million poisonings and 300,000 human deaths per year. Therefore, it is considered to be a global clinical problem. A range of nerve and muscular disorders are associated with acute or chronic exposure to OP compounds (see Ref. [13] and references within for details). The use of atropine and oximes, used for clinical management of OP poisoning, is generally inefficient (fatality rate of up to 40%) and ineffective in preventing post-poisoning neurocognitive dysfunctions. It is, therefore, imperative to develop an efficient therapy for patients with severe OP poisoning. In recent years, some bacterial enzymes capable of OP degradation, isolated

from soils, have shown promise for conferring prophylactic and therapeutic protection against OP compounds in mammals [14].

There is no doubt that environmental microbes are the major source of novel natural products because they are the most diverse (both structurally and metabolically) and abundant group of organisms and account for 60% of the Earth's biomass [15]. Current estimates suggest that soils provide habitat for approximately 4–5 $\times$ 10^{30} microbial cells, ten times more than the seas [16]. Because of their metabolic diversity, environmental microbes are the source of industrial products worth hundreds of billions of dollars [15]. Thus, understanding community and functional diversity is essential for exploiting the potential of microbes as a source of new drugs. Obtaining this information has been difficult owing to our inability to grow microorganisms under standard laboratory conditions (>99% of environmental microbes are unculturable), limiting the discovery of new products [1]. However, the ongoing revolution in genomic methodologies provides an opportunity to overcome this. Recent availability of genomic data suggests the potential of discovering new products from environmental microbes is hugely underestimated [3].

Metagenomics as an approach for new drug discovery from the uncultivable microbes

Metagenomics is being used more widely in an effort to explore and exploit the largely untapped reservoir of uncultured microbial genomes from natural environments. This approach to drug discovery involves the direct extraction of total genetic materials from environmental samples without the need for culturing. The genetic material is then ligated into vectors and transferred into new host cells (mainly *Escherichia coli*) to generate metagenomic libraries. These libraries can then be used to study structural and functional diversity by DNA sequencing or to search for new products based on sequence mining or functional expression. It is important, however, that the samples for metagenomics should be targeted from particular sites to increase the success rate because metabolites discovered by this approach depend directly on microorganisms present in the samples. For example, if the goal is an anti-tumour drug, screening of metagenomes from marine systems should be targeted because terrestrial samples show a low anti-tumour potential (0.01%) compared to those from marine systems (1%). Similarly, if a drug for detoxification is the target, samples with prior exposure to such toxicant might substantially increase the hit rate [5,13].

Extracting environmental genomes

Metagenomes have been extracted and expressed from several environments including marine, sediments, fresh water, soil and the human gut. The initial extraction of genomic DNA from such environmental samples is of fundamental importance if representative clone libraries and sequence information are to be obtained [17]. High-quality DNA that has not been fragmented is needed which requires less vigorous methods than those applied to diversity studies (e.g. bead beating). High-molecular-weight DNA is desirable because biosynthetic gene clusters encoding antimicrobial and other compounds of industrial interest typically range from 30 to 100 Kb. Thus, the recovery of high-molecular-weight DNA and subsequent insertion of large inserts is the most efficient method for identifying new enzymes [18]. Two methods that are routinely used include a direct lysis of environmental samples using freezing and thawing [19] and an indirect fractionation method whereby bacterial cells are isolated and purified from the environmental sample before lysis [20,18]. The direct lysis method is more commonly used because it is rapid and efficient and can confer high DNA yields [21]; however, this approach extracts total genomic material and is unable to discriminate between material of bacterial, archaeal or eukaryotic origin, and sorting redundant information can reduce the efficiency of screening. In addition, the co-extraction of undesirable compounds (e.g. humic substances) might result in some inhibition with downstream processing [22] using this direct lysis approach. Purification of the lysed product (e.g. electrophoresis) can be used to purify the DNA but might result in reduced yields. Alternatively, an indirect fractionation method can be used, whereby bacterial cells are separated from their environmental matrix before lysis [23]. Although more time consuming, this approach avoids problems associated with interfering contaminants (e.g. humic substances) and increases the efficiency of subsequent screening. A further problem is that extraction of total DNA can lead to an over-representation of a limited number of dominant genomes. This can be overcome to some extent through the use of cesium chloride gradient centrifugation to separate genomes on the basis of their GC content, DNA disassociation approaches or subtraction hybridization [24].

In some circumstances, such as low-nutrient environments (e.g. recent deglacial areas and rock), it might be problematic to obtain sufficient DNA for cloning. For such cases, whole-genome amplification can be used. Although this method is based on the use of short oligonucleotides that randomly amplify genomic DNA, it remains subject to the potential biases of PCR-based approaches [25]. Another approach is the use of multiple displacement amplification, which uses Ø29 DNA polymerase and a constant incubation temperature [26] to increase the amount of template DNA. Although multiple displacement amplification products are generally larger than those achieved from conventional whole-genome amplification approaches, they remain subject to amplification bias and the creation of chimeric sequences. It remains clear that better extraction efficiencies need to be developed for the recovery of good-quality DNA from difficult samples (e.g. highly humic soils or low template concentration), and some novel extraction methods are being explored (e.g. Ref. [27]).

Construction of metagenomic DNA libraries

A variety of vectors have been used in constructing metagenomic libraries. For small fragments (0.5–5 kbp), typically, plasmid-based vectors are cloned into host cells, most commonly *E. coli*. These vectors are unsuitable for large inserts, however, so to target entire functional operons or complete metabolic libraries, it is desirable to clone larger fragments (>35 kbp) in cosmid or fosmid (both 35–45 kbp) vectors or to use bacterial artificial chromosomes (typically 100 kbp). Cosmid and bacterial artificial chromosome vectors are commonly used in metagenomics [28–30]; however, because cosmid libraries can be unstable in maintaining prokaryotic or eukaryotic inserts [31], fosmid vectors are sometimes preferred [32].

Expression vectors of metagenomic libraries

E. coli is a convenient expression host because of the relative ease of genetically manipulating it, and it is commonly used in industrial fermentations. *E. coli* has been used successfully for expression of gene clusters involved in hydroxybutyrate utilization, carbonyl production and antimicrobial production [33–36], among others. However, the use of *E. coli* as an expression host is constrained in some cases in which high GC content in the inserts disrupts transcription in this low-GC-content host [37], notably in those constructs involved in antibiotic production. Ideally, the insert DNA and expression host should have similar GC content for successful expression. Alternative hosts include *Streptomyces* spp., *Pseudomonas* spp. and *Bacillus* spp. (e.g. Refs. [38–41]). Streptomycetes are particularly good hosts for the detection of secondary metabolites because their GC-rich genomes support a large proportion of biosynthetic genes that synthesize a considerable amount of secondary metabolites and, thus, provide a host capable of supplying precursors without the need for further genetic manipulation [3,41]. However, as a trade-off, these alternative hosts are often slow growing compared to *E. coli*, which might constrain their industrial efficiency. Attempts to improve the use of *E. coli* as a cloning and expression host might hold promise for more efficient expression [42,43]; therefore, further improvements in expression vectors and expression in surrogate organisms need to be tackled on a priority basis to increase the hit rate for new products from the metagenomic libraries.

Screening of metagenomic libraries

The two approaches for screening of metagenomic libraries to identify novel compounds are sequence mining and functional expression. Sequence mining includes sequencing of inserts in the whole metagenomic library and is advantageous in that it does not rely on heterologous expression. Development of automated and high-throughput sequencing technologies and bioinformatics has enabled sequencing of entire clone libraries in short time periods at reasonable cost. When the entire metagenome is available, searches for particular functions or proteins can be targeted by mining the sequence data. Once putative homologous sequences are found, exact sequence determination and expression of genes can be achieved by PCR amplification and expression in surrogate organisms. This approach has enhanced the discovery of several novel products including antibiotic and other therapeutic agents via detection of open reading frames with sequence homology to known functions [44]. Several multigenic pathways responsible for synthesis of the antibiotic violacein have been successfully isolated using this approach [45]. Despite this, direct sequencing of clone libraries remains costly because, generally, many clones need to be screened [46]. Current estimates suggest that to access all genomes within environmental samples, two million clones would be needed [47]. Furthermore, because the acquisition of complete gene clusters is not guaranteed, problems with downstream folding and expression may be encountered. Although this approach is not dependent on expression, it is unlikely to discover novel compounds because it relies on the design of probes based on available databases derived from already known genes. In addition, it is not able to identify novel genes and/or products that have the same function but a different sequence to known genes.

A more favourable approach is to use functional-based screening of clone libraries for the detection and expression of genes of interest because of the potential for discovery of novel products and direct industrial application. The greater success of functional-based screening for the identification of novel enzymes compared to sequencing approaches is illustrated in a recent review by Ferrer et al. [48]. Functional-based screening relies on detection of enzymes and encompasses simple colorimetric or other plate-based reactions, mediated by the desired products or protein, as well as more complicated screening whereby catalytic activity is identified and proteins are subsequently isolated (e.g. Ref. [49]). Functional screening of metagenomic libraries has been used successfully for the discovery of many industrially important compounds – such as antibiotics, esterases, lipases and oxidoreductases [27,29,36,49,50–52] – and for the identification of medically important genes involved in anti-infection, anti-proliferative and anti-inflammatory responses.

The use of metagenomic approaches for the isolation of novel products has been greatly accelerated in the past few years as sequencing technology has become faster and cheaper [1]. For example, several polyketides (with known activity as antibiotics, anti-tumour and immunosuppressants) were obtained from metagenomic libraries constructed from sponges, which are considered to be an important source of natural products [48], and the anti-tumour compounds pederin and bryostatin were isolated and characterized using a metagenomic approach. Pederin was isolated from the bacterial symbiont of rove beetle (*Paederus* sp.), and bryostatin was isolated from the marine protozoan *Begul neritina* using metagenomic techniques [3]. Similarly, indirubin, violacein and deoxyviolacein, and turbomycin A and B (all with high drug potential activity) were isolated from soil metagenomes. It is believed, however, that the hit rate of enzymatic activities ranges from 0.0001% to 0.2% and that of antimicrobial activities ranges from 0.009% to 0.03% [5]. The low hit rates are, in part, due to the barriers involved in logistics needed for screening the large number of sequences and problems associated with heterologous expression [53].

In recent years, there have been major breakthroughs in functional screening technologies. These include substrate-induced gene-expression screening, recently proposed by Uchiyama *et al.* [54], which offers an alternative screening strategy for the identification of clones displaying catabolic gene-expression. This approach is potentially more efficient and economical than sequence-based and other function-based screening [54]. A limitation, however, is that it cannot be applied to metagenomic libraries with large inserts [55]. A similar approach is to use genetic traps, whereby target genes are inserted between transcriptional regulators (e.g. downstream from a *luxR* promoter and upstream from a *gfp* promoter) that enable fluorescent cells containing the gene of interest to be captured by fluorescence-activated or inactivated cell sorting [56]. A further improvement is the development of laser-based technology, which claims to screen for up to one billion clones per day [48,57]. This still needs to be tested in various laboratories, however, and cost might be a barrier for many small laboratories. Therefore, further improvement in screening technologies will enable small companies to access products from the uncultivable microbes.

Because a large number of clones (up to two million) need to be screened to ensure that most genomes from an environmental sample are represented in clone libraries, this approach is considered expensive, time consuming and beyond the ability of small companies to exploit metagenomes for drug discovery. Genetic material enrichment approaches before cloning, however, may be used to enhance screening hit rates and reduce the number of clones to be screened. Enrichment approaches such as stable-isotope probing [58] or bromodeoxyuridine (BrdU) labelling [59] target metabolically active microbial populations and, thus, facilitate the efficiency of screening metagenomic libraries for genes of interest. Stable-isotope probing is used to selectively enrich genes of interest (primarily those involved in ^{13}C metabolism). Environmental samples are first incubated with a stable-isotope compound (primarily ^{13}C, although ^{15}N and ^{18}O have also been used); microbes that utilize this compound incorporate the stable-isotope label into their cellular material, including DNA and RNA. Total genetic material is extracted from the enriched environmental sample and the labelled DNA/RNA is subsequently fractionated from unlabelled DNA/RNA via ultracentrifugation. Heavy (isotope-labelled) DNA/RNA is then used to construct metagenomic libraries. Because this approach mainly includes cloning of DNA from a small population that utilized the substrate, the number of clones needing to be screened for that particular function is small. Similarly, BrdU is incorporated into replicating DNA and when combined with the addition of a substrate can be used to selectively enrich for particular functions. BrdU-labelled DNA is then separated from non-labelled DNA through the use of immunocapture bead technology. Although such approaches have some bias towards fast-growing species (thus potentially reducing the discovery of novel genes from numerically small and slow-growing microbes [60,61]), they can expedite the process of new product discovery as a result of reduced cost, logistics and time required for screening. These approaches have been used successfully to obtain a new gene system for novel biphenyl-degrading enzyme [62] and will be particularly suitable for isolation of the agents of medical bioremediation, protein therapies and treatment for chemical poisoning.

Other approaches

Whole-genome sequencing of several organisms including bacteria provides a wealth of information on new drugs. Whole-genome sequencing of bacteria has shown that the number of biosynthetic gene clusters outnumber – by some distance – the number of currently known metabolites [3]. So far, a little over 1000 microbial genomes have been fully sequenced (source: NCBI http://www.ncbi.nlm.nih.gov/). Genome sequence mining of these cultivable bacteria has already led to the discovery of new drugs. For example, new anti-cancer drugs (diazepinomicin and Eco-7942) were discovered from the genome mining of *Micromonospora* and *Streptomyces* spp. Similarly, several anti-helmintic drugs (E-637 and E492) and antibacterial drugs were reported from the genome mining of *Streptomyces* and *Amycolatopsis* spp. Currently, less than 1% of microbes can be cultured in laboratory; therefore, it is imperative to the opening of new frontiers to isolate products from the uncultivable microbes. One such approach can be direct sequencing of metagenomic DNA without the need to construct a clone library. Such an approach enables one to predict the potential of target molecule presence in the sample, which can be later harvested by PCR and heterologous expression. This approach has been further developed, so the whole-genome of a particular functional species can be sequenced without culturing. For example, Kalyuzhnaya et al. [62] used stable-isotope probing and whole-genome shotgun sequencing to recover functional metagenomes of methylotrophic microbes, resulting in the reconstruction of substrate breakdown and near-complete genome sequencing of a few microbes responsible for methylotrophic metabolism [63].

As these approaches mature and become available to smaller industries in the future, it will aid the continued discovery of new drugs and ultimately enable realization of the full potential of microbial drug discovery. Another interesting but controversial approach is reactome arrays, described recently by Beloqui *et al.* [64]. This approach reports direct isolation of enzymes from environmental samples without the need for culturing or creating metagenomic libraries. Here, the substrate is linked with a dye and degradation of compound is recorded with concomitant release of the dye signal. The enzyme responsible for degradation is captured on nanoparticles coated with cognate metabolite. Metagenomics and reactome array can be used as complimentary approaches to discover novel products: for example, where heterologous expression of gene is a problem but metagenomic sequence mining suggests the presence of novel microbial products in a particular environmental sample, reactome array can be used to directly capture that particular protein. Conversely, if reactome array suggests the presence of particular protein in an environmental sample but quantity of product is an issue, protein sequencing can be done to discover the amino acid composition, which can be used to design PCR primers for amplification of the target gene for subsequent cloning and expression in a surrogate host. If repeatable in other laboratories, with further improvement in efficiency, this approach could represent a real boost in protein therapies, medical bioremediation and treatment of chemical poisoning.

Integration with other emerging technologies

Obtaining proteins with therapeutic activities is only the first step. Changes in protein activities and stability will be required to achieve efficiency and economic viability [65]. For the successful use of microbial enzymes or proteins in human treatment, however, two main problems need to be overcome: immunological reactions owing to the injection of exogenous protein and blood protease attacks on injected protein [14]. To address these problems, microbial proteins need to be formulated in a carrier that provides a protective environment for the enzyme while remaining permeable to target molecules for intravenous treatment, avoiding immunological reactions [13]. Recent advancements in nanotechnology offer a promising tool to overcome both problems [15]. The nanotechnology approach and its applications have been reviewed extensively [66,67] and, therefore, are not discussed in detail in this article. In brief, biomedical applications of nanotechnology include adsorption, covalent bonding, entrapping of microbial products on nanostructures (nanoparticle, nanofibre, nanocrystal, and so on) to provide stability of the products, and targeted delivery to cells and tissues [68]. Because of their small size (1–100 nm), nanomaterials possess unique magnetic,

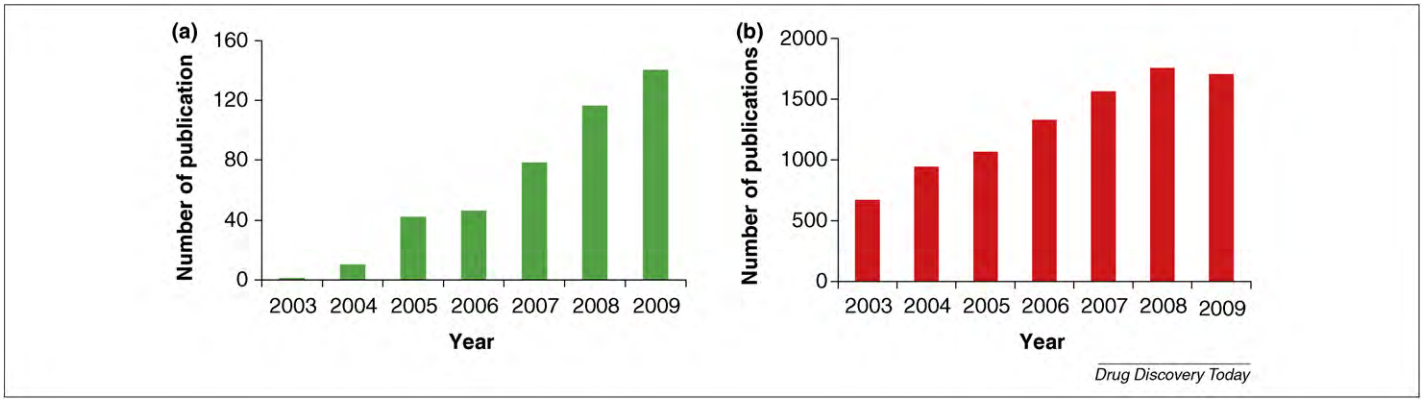

FIGURE 1

A graphical presentation of a literature survey using the ISI web knowledge tool on the use of metagenomics and nanotechnology. It indicates that the use of **(a)** metagenomics and **(b)** nanotechnology has accelerated in recent years.

thermodynamic and catalytic properties, which are exploited by the pharmaceutical industry for targeted delivery and stability of medicines [66]. Entrapment of microbial products in nanostructures has already been reported to protect from protease attack and minimize immunological responses [14]. Several nano-formulated drugs are already being prescribed or under clinical trials (e.g. Rapamune®, an immunosuppressant, and Emend®, a medicine for emesis are used globally [67]). Further improvement in nano-

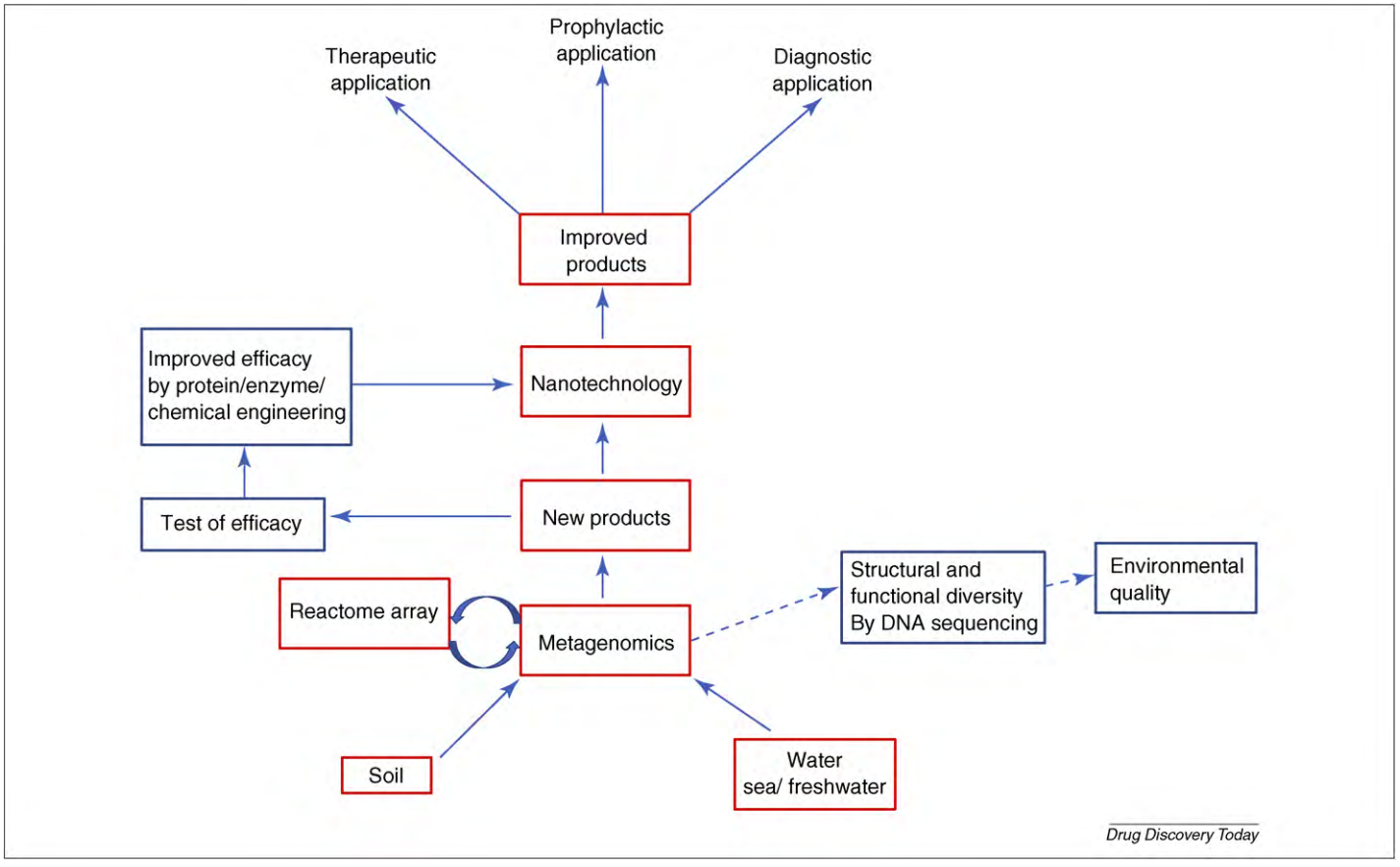

FIGURE 2

A flow-chart to demonstrating an integrated approach to combine metagenomics, reactome array, protein and enzyme engineering and nanotechnology to tap the potential of microbial products for drug discovery. Metagenomics can be used to produce various novel and more efficient microbial products which can be further improved, if needed by protein and enzyme engineering. Metagenomics and reactome array can be used as complimentary approaches to discover novel products. For example, where heterologous expression of gene is a problem but metagenomic sequence mining suggests the presence of novel microbial products in a particular environmental sample, reactome array can be used directly to capture that particular protein. By contrast, if reactome array suggests the presence of particular protein in a environmental sample but quantity of product is an issue, protein sequencing can be done to know the amino acid composition which can be used to design PCR primers for amplification of the target gene for subsequent cloning and expression in surrogate host. The microbial products can then be modulated by nanotechnology which can increase their activity and stability, and protect from the protein attacks and immunological responses.

technology for its integration with metagenomic products will increase the chance of exploitation of the number of microbial products obtained from cultured or metagenomic approaches.

Future perspectives

Metagenomics and nanotechnology have started to mature in terms of application and new drug discovery. A literature survey indicates that the use of metagenomics and nanotechnology have accelerated in recent years (Figure 1). A survey using ISI web knowledge suggests that up to 140 papers on 'metagenomic approach' and 1704 papers on 'nanotechnology' have been published in 2009 alone. When these two terms are combined in the search, however, only one published paper can be found [69], which concerns the use of nanotechnology for the development of sequencing machines rather than their combined use for advancing biotechnological applications. This finding clearly indicates that these two techniques have been used in isolation in all fields, including drug discovery. Nonetheless, the health industries have already started benefiting from the emerging fields of metagenomics and nanotechnology, but we are not even near to realization of their potential in drug discovery. Microbes and their products are a huge potential resource (and probably the largest source of new drugs), but our ability to extract these products is largely limited by our inability to grow the vast majority of environmental microbes. The emergence of metagenomics and, more recently, reactome arrays provide for the first time the ability to extract these drugs without culturing the microbes that produce them. Metagenomics as a technology is maturing, and already, several microbial products with therapeutic properties have been isolated using the methodology. It is likely that our ability to exploit this approach will multiply several-fold if we can find solutions to two obstacles: efficient heterologous expression and high-throughput functional screening. Considerable advancements have been made on both fronts in recent years, but further improvement is needed to make expression and screening cheaper, less labour intensive and more efficient. Reactome arrays promise a huge potential to extract proteins directly from uncultured microbes, and if they can be optimized to extract low amounts of proteins or secondary metabolites produced in environmental samples, it will be a huge advancement in the search for natural products with therapeutic usages.

The real potential of these approaches, however, can only be realized if we can integrate all approaches from the conceptual stage (Figure 2). Several novel products (especially proteins) obtained from the metagenomic and reactome array approaches will still need modification to achieve stability and efficiency of the product. This can be achieved by protein and enzyme engineering. Further improvement using nanotechnology formulation will enhance activity by helping in targeted delivery, protection against protease attacks and minimum or no immunological responses. Therefore, to harness the real potential of environmental microbes in drug discovery, the use of an interdisciplinary approach combining genomics (metagenomics and functional genomics), protein and enzyme engineering, and nanotechnology is needed. A process is already underway to achieve this, but a concerted and sustained effort from the stakeholders (industries, academics, funding agencies and regulatory agencies) will expedite the process and might open a door to numerous new therapeutic products in the near future.

References

1 Li, J.W-H. and Vederas, J.C. (2009) Drug discovery and natural products: end of an era or an endless frontier? *Science* 325, 161–165
2 Koehn, F.E. and Carter, G.T. (2005) The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* 4, 206–220
3 Van Lanen, S.G. and Shen, B. (2006) Microbial genomics for the improvement of natural product discovery. *Curr. Opin. Microbiol.* 9, 252–260
4 Demain, A.L. (2009) Antibiotics: natural products essential for human health. *Med. Res. Rev.* 29, 821–841
5 Pettit, R.K. (2004) Soil DNA libraries for anticancer drug discovery. *Cancer Chemother. Pharmacol.* 54, 1–6
6 Hranueli, D. *et al.* (2005) Plasticity of the *Streptomyces* genome – evolution and engineering of new antibiotics. *Curr. Med. Chem.* 12, 1697–1704
7 Leader, B. *et al.* (2008) Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* 7, 21–39
8 Reichert, J.M. (2003) Trends in development and approval times for new therapeutics in the United States. *Nat. Rev. Drug Discov.* 2, 695–702
9 Mathieu, J.M. *et al.* (2009) Medical bioremediation of age-related diseases. *Microb. Cell Fact.* 8, 21
10 Mathieu, J. *et al.* (2008) Microbial degradation of 7-ketocholesterol. *Biodegradation* 19, 807–813
11 Casserly, I. and Topol, E. (2004) Convergence of atherosclerosis and Alzheimer's disease: inflammation, cholesterol, and misfolded proteins. *Lancet* 363, 1139–1146
12 Zhao, H. *et al.* (2008) Amyloid-beta peptide degradation in cell cultures by mycoplasma contaminants. *BMC Res. Notes* 1, 38
13 Singh, B.K. (2009) Organophosphorus degrading bacteria: ecology and industrial application. *Nat. Rev. Microbiol.* 7, 156–164
14 Sogorb, M.A. *et al.* (2004) Future application of phosphotriesterases in the prophylaxis and treatment of organophosphorus insecticide and nerve agent poisonings. *Toxicol. Lett.* 151, 219–233
15 Singh, B.K. *et al.* (2009) Soil metagenomics. *Nat. Rev. Microbiol.* 7 10.1038/nrmicro2119-c1.
16 Curtis, T.P. and Sloan, W.T. (2005) Exploring microbial diversity – a vast below. *Science* 309, 1331–1333
17 Bertrand, H. *et al.* (2005) High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J. Microbiol. Methods* 62, 1–11
18 Lefevre, F. *et al.* (2007) Environmental metagenomics: innovative resourse for industrial biocatalysis. *Biocat. Biotransform Rev.* 25, 242–250
19 Zhou, J. *et al.* (1996) DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–322
20 Courtois, S. *et al.* (2001) Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ. Microbiol.* 3, 431–439
21 Roose-Amsaleg, C.L. *et al.* (2001) Extraction and purification of microbial DNA from soil and sediment samples. *Appl. Soil Ecol.* 18, 47–60
22 Tsai, Y.L. and Olson, B.H. (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl. Environ. Microbiol.* 2292–2295
23 Robe, P. *et al.* (2003) Extraction of DNA from soil. *Eur. J. Soil Biol.* 39, 183–190
24 Cowan, D. *et al.* (2005) Metagenomics gene discovery: past present and future. *Trends Biotechnol.* 23, 321–329
25 Bodelier, P.L.E. *et al.* (2009) Whole community genome amplification (WCGA) leads to compositional bias in methane-oxidizing communities as assessed by *pmoA*-based microarray analyses and QPCR. *Environ. Microbiol. Rep.* 1, 434–441
26 Abulencia, C.B. *et al.* (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl. Environ. Microbiol.* 72, 3291–3301
27 Yun, J. *et al.* (2004) Characterization of a novel amylolytic enzyme encoded by a gene from a solid-derived metagenomic library. *Appl. Environ. Microbiol.* 70, 7229–7235
28 Rondon, M.R. *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66, 2541–2547

29 Courtois, S. *et al.* (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* 69, 49–55
30 Daniel, R. (2004) The soil metagenome—a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15, 199–204
31 Beja, O. (2004) To BAC or not to BAC: marine ecogenomics. *Curr. Opin. Biotechnol.* 15, 187–190
32 Voget, S. *et al.* (2003) Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* 69, 6235–6242
33 Henne, A. *et al.* (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* 65, 3901–3907
34 Knietsch, A. *et al.* (2003) Metagenomes of complex microbial consortia derived from different soils as sources of novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 5, 46–56
35 MacNeil, I.A. *et al.* (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* 3, 301–308
36 Gillespie, D.E. *et al.* (2002) Isolation of antibiotics tubomycon A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* 68, 4301–4306
37 Richard, C.L. *et al.* (2004) *Rhodobacter capsulatus nifH*A1 promoter: high-GC-10 regions in high-GC bacteria and the basis for their transcription. *J. Bacteriol.* 186, 740–749
38 Wang, G.Y. *et al.* (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* 2, 2401–2404
39 Martinez, A. *et al.* (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* 70, 2452–2463
40 Rondon, M.R. *et al.* (1999) The earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol.* 17, 403–409
41 Lorenz, P. and Eck, J. (2005) Metagenomics and industrial applications. *Nat. Rev. Microbiol.* 3, 510–516
42 Pfeifer, B.A. *et al.* (2001) Biosynthesis of complex polyketides in metabolically engineered strain of *E. coli*. *Science* 291, 1790–1792
43 Boddy, C.N. *et al.* (2004) Precursor-directed biosynthesis of epothilone in *Escherichia coli*. *J. Am. Chem. Soc.* 126, 7436–7437
44 Yang, Y. *et al.* (2009) Target discovery from data mining approaches. *Drug Discov. Today* 14, 147–154
45 Brady, S.F. *et al.* (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* 3, 1981–1984
46 Henne, A. *et al.* (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 66, 3113–3116
47 Ginolhac, A. *et al.* (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* 70, 5522–5527
48 Ferrer, M. *et al.* (2009) Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* 16, 109–123
49 Ferrer, M. *et al.* (2005) Novel microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem. Biol.* 12, 895–904
50 Richardson, T.H. *et al.* (2002) A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. *J. Biol. Chem.* 277, 26501–26507
51 Brennan, Y. *et al.* (2004) Unusual microbial xylanases from insect guts. *Appl. Environ. Microbiol.* 70, 3609–3617
52 Knietsch, A. *et al.* (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 69, 1408–1416
53 Handelsman, J. *et al.* (2002) Cloning the metagenome: culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods Microbiol.* 33, 241–255
54 Uchiyama, T. *et al.* (2005) Substrate induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* 23, 88–93
55 Yun, J. and Ryu, S. (2005) Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microb. Cell Fact.* 4, 8
56 Williamson, L.L. *et al.* (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl. Environ. Microbiol.* 71, 6335–6344
57 Podar, M. *et al.* (2007) Targeted access to the genomes of low-abundance organisms in complexed microbial communities. *Appl. Environ. Microbiol.* 73, 3205–3214
58 Neufeld, J.D. *et al.* (2007) DNA stable-isotope probing. *Nature Prot.* 2, 860–866
59 Urbach, E. *et al.* (1999) Immunochemical detection and isolation of DNA from metabollically active bacteria. *Appl. Environ. Microbiol.* 65, 1207–1213
60 Neufeld, J.D. *et al.* (2007) Methodological considerations for the use of stable isotope probing in microbial ecology. *Microb. Ecol.* 53, 435–442
61 Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* 14, 303–310
62 Sul, W.J. *et al.* (2009) DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl contaminated river sediment. *Appl. Environ. Microbiol.* 75, 5501–5506
63 Kalyuzhnaya, M.G. *et al.* (2008) High resolution metagenomics targets specific functional types in complex microbial communities. *Nat. Biotechnol.* 26, 1029–1034
64 Beloqui, A. *et al.* (2009) Reactome array: forging a link between metabolome and genome. *Science* 326, 252–257
65 Singh, B.K. (2010) Exploring microbial diversity for biotechnology: the way forward. *Trends Biotechnol.* 28, 111–116
66 Sanvicens, N. and Marco, M.P. (2008) Multifunctional nanoparticles – properties and prospects for their use in human medicine. *Trends Biotechnol.* 26, 425–433
67 Junghanns, J.U.A.H. and Muller, R.H. (2008) Nanocrystal technology, drug delivery and clinical applications. *Int. J. Nanomed.* 3, 295–309
68 Moore, J.A. (2006) Small science, big challenge. *Nature* 442, 747
69 Hudson, M.E. (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resour.* 8, 3–17

Further reading

70 Chin, Y.-W. *et al.* (2006) Drug discovery from natural sources. *AAPS J.* 8, E239–E253
71 Lam, K.S. (2007) New aspects of natural products in drug discovery. *Trends Microbiol.* 15, 279–289