1 Introduction

Microorganisms have a long history as a resource for novel enzymes, biocatalysts and biologically active compounds [1]. The concept of microbial diversity has dramatically expanded within the past decade. It has been known for a long time that many prokaryotic species are difficult to cultivate in the laboratory because of specialized growth requirements. If one compares direct microscopic cell counts after 4’,-6-diamidino-2-phenylindole (DAPI) staining of bacterial DNA with the number of microbial colonies growing on nutrient agar, it appears that in natural samples fewer than one cell in a thousand produces a colony, a phenomenon known as the “great plate count anomaly” [2]. According to Amann and colleagues [3] only a minority of the microorganisms living in any given habitat are cultivable. Amann reported that 0.001–0.1% of the microorganisms in seawater, 0.25% in fresh water, 0.25% in sediments and only 0.3% of soil microorganisms were found to be cultivable. Conversely, typically 99% of the microbial diversity in any habitat is not accessible using classical microbiological cultivation technologies. Through direct isolation and cloning of the complete genomes of all microorganisms living in a defined habitat (Fig. 1), it is possible to access the total genetic information, the so-called “Metagenome” without any need to isolate and cultivate any cells [4]. Metagenomics, the application of the entire genomics technology suite to metagenomic DNA, has the potential to substantially impact industrial production [5]. The enormous dimensions of the biological and consequently molecular diversity, was sketched out in actual numbers by Torsvik et al. [6] and Venter et al. [7]. A pristine soil sample may contain in the order of $10^4$ different bacterial species. More than one million novel open reading frames, many of which encode putative enzymes, were identified in a single effort that sampled marine prokaryotic plankton retrieved from the Sargasso Sea, thus giving access to a nearly inexhaustible genetic resource for biomolecules of potential utility in a variety of industries. Habitats targeted in metagenomics can be as diverse, as summarized in Fig. 2, and may include termite guts, the rumen of ruminants like deer or cows, soil samples from deserts and glaciers, as well as extreme habitats like geysers, open sea water or deep sea hydrothermal vents. It is to be expected that the genomic compositions of the microbial populations in these habitats will differ from each other, and with it the biocatalyst and biomolecule composition of the respective metagenome libraries. Industrial metagenomics primarily focuses on prokaryotes, as their genomes can easily be targeted by the functional screening tools available today, and because it is assumed that
maximum biodiversity is to be found in the bacterial lineages, as reflected in the published literature [8–10].

2 Attractive industrial biocatalysts from metagenome

Enzymes are used in a wide range of applications and industries [11]. They are required in small quantities to synthesize kilogram amounts of a chiral synthon that is used as a building block to produce highly active pharmaceuticals, or at a kiloton/year scale to be added as active ingredients in bulk products such as high-performance laundry detergents [12]. Their natural versatility allows their use as much in processes to degrade natural polymers including starch, cellulose and proteins, as well as for the regio- or enantioselective synthesis of asymmetric chemicals, although there are many more applications.

Whereas the food, feed and detergent industries typically concentrate on a limited number of enzyme reactions and substrates, the chemical and pharmaceutical industries deal with thousands of chemically and structurally diverse molecules, and production of each of these requires individual enzymatic solutions. Consequently, owing to a wealth of potentially useful biocatalysts, microbial resources are increasingly popular with the chemical industries, and are viewed as indispensable for the modern organic chemist [13]. With numerous successfully implemented processes running worldwide [14, 15], and the number of industrialized biotransformations having doubled every decade since 1960 [16], it is estimated that in 10% of processes biocatalysis may provide a superior synthetic solution over classical chemistry [17]. In fact, the availability of an appropriate biocatalyst is now thought to be a limiting factor for any biotransformation process [14].

Thus, there is clearly ample demand for novel enzymes and biocatalysts, and metagenomics is currently thought to be one of the most likely technologies to provide the candidate molecules required [18, 19].

What are industry’s key requirements for an interesting biocatalyst? For any industrial application enzymes need to function sufficiently well according to several application-specific performance parameters, e.g., activity, stability, specificity and efficiency. The additional value that metagenomics and its harvest of biomolecules has to offer industry is specific for every industry addressed:

1. Novelty: For industries that produce bulk commodities, such as high-performance detergent formulas, a single enzyme backbone with superior functionality that has an entirely new sequence would be very attractive, to avoid competing intellectual property rights. The relevance of this issue is illustrated for detergent proteases (subtilisins) by the fact that substitutions at nearly every position in the mature 275 amino acid Bacillus protease Novo type (BPN’). subtilisin have been claimed in patents [12].

2. Maximum diversity: The pharmaceutical and supporting fine chemicals industries often seek entire sets of multiple diverse biocatalysts to build in-house toolboxes for biotransformations [20]. These toolboxes need to be easily accessible to meet the timelines of a stringent biosynthetic feasibility evaluation in competition with traditional synthetic chemistry (see below).
Elusive metabolites: Many pharmacologically active secondary metabolites are produced by bacteria that live in complex consortia, or by bacteria that inhabit niches that are difficult to reconstitute in vitro [21]. Therefore, the cloning and heterologous expression of biosynthetic genes that encode secondary metabolites (usually present as gene clusters) is the most straight forward and reproducible method of accessing their biosynthetic potential; a path followed by several biopharmaceutical companies (see below).

3 The market opportunity for the chemical industry: implementation of white biotechnology

White, or industrial biotechnology has the potential to greatly impact industrial production processes on a global scale. Major long-term applications of white biotechnology will be replacing fossil fuels with biofuels (biomass conversion), replacing and supplementing conventional chemical processes with bioprocesses (including metabolic engineering) and creating new high-value bioproducts including nutraceuticals, performance chemicals and bioactives mostly with the help of novel biocatalysts or enzymes. That environmentally sound, commercially viable biotechnological processes can take their respected place in a global industrial environment has been acknowledged for several years [22]. Companies from Europe, Canada, Japan, South Africa and the USA reported on their experiences in processes as diverse as the production of acrylamide (Mitsubishi Rayon, Japan) and the use of enzymes in oil-well completion (British Petroleum Exploration, UK). Right now this movement towards implementing sustainable technologies and processes is gaining momentum in Europe in particular. Besides the involvement of the food and feed, detergent and the politically heftily promoted biofuel industries, it is the globally operating chemical and pharmaceutical industries that are active players in white biotechnology. As industries face increasing low-cost competition, particularly from East Asia, and political pressure to reduce their environmental impact and resource consumption to improve sustainability, it is felt that there is a strong need for smart and innovative technologies, processes and products to remain competitive.

The McKinsey consultancy projects that by 2010 biotechnology could be applied in the production of between 10% and 20% of all chemicals sold (amounting to a value between US$ 160 billion and US$ 300 billion) and that up to 60% of all fine chemicals (medium volume products used as intermediates in the manufacturing of products such as pharmaceuticals, flavors and fragrances, agro-chemicals or detergents) might be produced using biotechnology [23]. Even for the traditional mainstay of the chemical industries, the polymer market (typical bulk
According to Business Communications Co. [24], the global industrial enzyme market was estimated at US$ 2 billion in 2004. The main profits were divided among technical enzymes with US$ 1050 million, including detergent, pulp and paper; food applications (US$ 720 million), and agriculture/feed (US$ 230 million). In the face of soaring energy costs, dwindling fossil resources, environmental pollution and a globalized economy, the large scale use of biotechnology instead of, and complementing, traditional industrial production processes, particularly in the chemical sector, is viewed as both an opportunity and a necessity. In the future novel biotechnological applications will boost the market for industrial enzymes. The authors of the BCC report expect the enzyme market to grow to some US$ 2.4 billion by 2009.

4 Screening for industrial enzymes and biocatalysts in metagenomes

After Torsvik’s report in 1980 on the extraction and digestion of genomic DNA from bacteria isolated from soil [25], and the publication of the concept of generating gene libraries directly from environmental DNA [26], it was only in the next decade that the first metagenome libraries were reported from genomic DNA fragments isolated directly from marine plankton [27] and from enriched consortia of cellulose digesters [28]. The work on the enriched consortia was particularly relevant in that it appears to describe the first successful expression-screening of metagenomic DNA. Once this was achieved, however, the utility of the metagenome approach for biotechnology was clear. Several companies, small and medium-sized enterprises as well as larger corporations are employing metagenomics to discover novel enzymes, among them Diversa (San Diego, CA, USA) [29, 30], Genencor (Palo Alto, CA, USA) [31], Henkel (Düsseldorf, Germany) [32], and Degussa (Düsseldorf, Germany) [33].

Both sequence-similarity-based screens, targeting new genes that were related to known references and activity-linked expression strategies were used by BRAIN AG (Zwingenberg, Germany) in industrial collaborations with Henkel targeting novel glycosyl hydrolases for use in laundry applications [32] and with Degussa to compile a collection of novel nitrile hydratases [34]. Although both sequence-similarity-based screening and activity-based screening may retrieve valuable catalysts from the metagenome (Fig. 3), the activity based approach holds more potential for identifying entirely novel active sequence space.

5 The quest for biocatalyst toolboxes

With metagenomics offering an unparalleled wealth of diverse biocatalysts with biotechnological potential, industries dealing with a multitude of substrates and processes will be attracted to harness this resource for different reasons.

Biocatalysis to achieve efficient chemo-, regio- and enantioselective conversion of precursors to high-purity products has been a long-time goal of the biotechnology industry. The key challenge for industry is to develop a process for the production of high-purity products, often with added value, that is both efficient and economically viable. The quest for biocatalyst toolboxes aims to address this challenge by identifying and developing new biocatalysts that can be used in industrial processes.
and high-value chemicals, pharmaceuticals or intermediates is increasingly commonplace in modern synthetic chemistry [13, 35]. In setting up a synthetic route for a desired fine or specialty chemical, particularly a synthon leading to an active pharmaceutical ingredient, very early feasibility studies determine which, if any, step may be carried out using isolated enzymes or whole cell biocatalysts. Here fast in-house access to validated proprietary resources and, in consequence, to proprietary applications are key to industrial success. The decision timelines particularly in early pharmaceutical development are in the order of weeks or months and typically substrates are top secret and must not be disclosed or physically passed on to third parties. Therefore, a multitude of analytical scale samples of diverse biocatalysts in a pre-characterized, ready-to-use format must be available on the shelf for testing.

In this respect, the requirements to be met by a collection of biocatalysts in the pharmaceutical and supporting industries are possibly the toughest in the business, although pressure and timelines in other industries do not seem to be significantly more relaxed.

A very different view on the utility of a collection of catalysts with relevant activity and substrate profiles may come from industries (e.g., the food, feed or detergent industries) with more limited numbers of substrates. Here it may be important to legally claim a certain functional sequence space, i.e., a set of similar but non-identical enzyme sequences with desired activities. This would generate a freedom to operate and at the same time hamper exploitation of such valuable biocatalysts by competitors.

If a certain type of substrate class and/or reaction chemistry defines the core expertise of a fine or specialty chemicals company, it makes sense to accumulate as many different reaction and substrate specificities in the form of diverse biocatalysts as reasonably possible. These can then be used to address recurring catalytic or chemosynthetic problems (e.g., selective reduction of prochiral ketones, oxygen-functionalization of aromatics or the hydrolysis of esters or amides for chiral resolution). The advantages of having a biocatalyst toolbox at hand have been realized by several industrial leaders who in some cases actively promote this fact [20]. Examples of biocatalyst toolboxes for industrial applications established by BRAIN are given in scientific and patent literature [32, 33, 36].

6 Searching for bioactives in metagenomes

Despite large increases in the R&D expenditure of big and medium-sized pharmaceutical companies, the number of pharmacologically active New Chemical Entities (NCEs) has not increased proportionally [37], and this is most notable in the paucity of new antibiotics. Big pharmaceutical companies currently find antibiotics economically un-attractive to develop for a variety of reasons when compared with drugs to treat long-term chronic conditions like obesity and high cholesterol [38]. This could have dramatic consequences for the global antibiotic resistance problem. Natural compounds and their derivatives will continue to have a significant role in drug discovery, as they have done in the past [39]—natural products constituted 63% of all newly approved anti-infectives between 1983 and 1994 [40]. Many natural products are bacterial and fungal secondary metabolites, but since most microorganisms cannot easily be cultivated, it is likely that many potentially active compounds have never been characterized. Metagenomics might have an invaluable role in the discovery process of new bioactive molecules [41, 42]. For this reason, metagenomes have attracted the attention of academia and specialized pharmaceutical companies [43–46].

7 Metagenomics – an industrial outlook

7.1 Novel biocatalysts and enzymes

As the excitement about the genetic access to boundless realms of microbial diversity slowly gives way to the routine of tapping into this diversity, the inevitable challenge of heterologous gene expression needs to be addressed to turn metagenomic technologies into commercial successes, particularly in applications where bulk enzyme or product quantities have to be produced at very competitive prices.

Metagenomics, together with in vitro evolution and high throughput screening technologies, provides industry with an unprecedented chance to bring biomolecules into industrial application. That this is truth and not fiction has been shown by Diversa, BRAIN and others. A full cycle from the discovery of novel molecular scaffolds from multiple resources, including metagenomes to the recombination by in vitro evolution technologies to generate an improved biocatalyst for a specific application has been realized, with production of alpha-amyloses for applications in the hot and acidic process of starch liquefaction [47], demonstrating the feasibility of the “ideal biocatalyst” concept. Once new genes are cloned and screened for activity the main stumbling block is the expression of sufficiently pure protein in useful amounts at reasonable costs. Clearly, cheap and efficient enzyme production in high-performance expression systems involving bacilli or filamentous fungi is a key factor for success, particularly when the enzyme functions as part of the final (bulk) product (e.g. in detergents). In the fine chemicals industry, this may be similar for bulk product synthesis. However, particularly in the pharmaceutical industry, time to market is decisive, and in these applications it may be even more important for a company to have a large collection of biochemically diverse catalysts at hand for
rapid testing, even if these molecules are not expressed in large amounts.

7.2 Novel bioactives
Besides the well-established metagenome approach for the identification of novel biocatalysts, its obvious potential for drug discovery has been realized by companies around the world. Whereas small to medium-sized enterprises have been recently engaged in metagenome research for the sake of anti-infectives discovery, other uses for bioactive molecules in cancer therapy and other medical fields will presumably lure companies currently engaged in red biotechnology into utilizing metagenomics [21].

8 References