Journal of Molecular Catalysis B: Enzymatic 68 (2011) 22–29

Contents lists available at ScienceDirect

Journal of Molecular Catalysis B: Enzymatic

journal homepage: www.elsevier.com/locate/molcatb

Review

Intragenomic enzyme complements

Gideon Grogan*

York Structural Biology Laboratory, Department of Chemistry, University of York, YO10 5YW York, United Kingdom

ARTICLE INFO

Article history:
Received 23 March 2010
Received in revised form
10 September 2010
Accepted 27 September 2010
Available online 28 October 2010

Keywords:
Biocatalysis
Enzyme discovery
Genomics
High-throughput
Oxidoreductase

ABSTRACT

Researchers in applied biocatalysis are now reaping the rewards of intensive effort and technological developments in the sequencing of the genomes of microbial and plant species. The genomic resource contains the sequences of millions of new genes with potential application in industrial biotechnology and includes families of enzymes within discrete genomes that potentially catalyze equivalent chemical reactions. One of the key emerging characteristics of these *intra*genomic complements of enzymes is the impressive breadth of catalytic diversity that is observed within them. This diversity may have been acquired either in order to combat the spectrum of metabolic challenges with which the organism may be presented in its natural environment or as part of the biosynthetic machinery evolved to produce a spectrum of secondary metabolites that will prove to be advantageous in establishing a niche. Attempts have been made to functionally characterize the intragenomic complements of enzyme families catalyzing diverse reactions including carbonyl reduction, ester hydrolysis and Baeyer–Villiger oxidation, in Gram-positive bacteria, yeasts, filamentous fungi and the plant *Arabidopsis thaliana*. These studies are beginning to describe in detail for the first time the impressive range of catalytic potential within single organisms for attributes such as substrate range, enantioselectivity or thermostability, each of which is of interest from an enzyme discovery perspective.

Contents

1. Introduction

At the time of writing, the Genomes OnLine Database (GOLD, http://genomesonline.org/) lists over eight thousand genome sequencing projects with over 1300 publicly-accessible completed genomes. It is noteworthy that approximately a quarter of these have been listed as having a 'funding relevance' in the Biotechnology Sector. The reasons for this are clear: Understanding the biochemistry of industrial organisms such as *Corynebacterium glutamicum* [1], *Aspergillus niger* [2], and *Streptomyces avermitilis* [3] will assist greatly in the improved exploitation of these organisms for the production of bulk biochemicals such as amino acids and citric acid, or fine chemicals such as antibiotics, respectively. The other beneficial consequence of such genome sequences is that they reveal multitudes of related genes encoding enzymes that are of particular interest in the area of applied biocatalysis, such as hydrolases (proteases, esterases, nitrilases, lipases), oxidoreductases (cytochromes P450, ketoreductases, flavin-dependent oxygenases) transferases (glycosyltransferases and aminotransferases) as well as open-reading frames that potentially encode enzyme chemistries that have not been previously described [4]. In this review, we explore the extent to which the *intra*genomic diversity of catalytic activity within a single species has been explored, and how this might impact on enzyme discovery for biocatalysis.

* Tel.: +44 1904 328256; fax: +44 1904 328266.
E-mail address: grogan@ysbl.york.ac.uk.

1381-1177/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molcatb.2010.09.012

Initial studies reveal that there is a surprising amount of diversity in sequence and catalytic properties within these small groups of enzymes, which constitute a readily accessible discrete pool of catalysts that are remarkably complementary in terms of reaction chemistry, substrate specificity, and regio- and enantioselectivity.

2. The genomes of 'biocatalytic' organisms and their enzyme complements

Over the last forty years, certain microbial genera have emerged as the most frequently encountered during screening and enrichment selection experiments conducted in an effort to identify and exploit a particular biocatalytic reaction. These organisms would be well-known to scientists practicing in the area of applied biocatalysis, and some representative strains were early targets for genome sequencing, including for example, strains of the bacteria *Pseudomonas* [5,6], *Bacillus* [7], *Rhodococcus* [8], *Nocardia* [9], *Arthrobacter* [10], various *Streptomyces* spp. [3,11–13] and also filamentous fungi, for which strains of *Aspergillus* [2,14,15] are the most notable. Strains from these genera have been applied as whole-cell biocatalysts to a host of chemical reactions ranging from ester, epoxide and nitrile hydrolysis to the hydroxylation of non-activated carbon centres and other oxidations. In some cases, protein purification, followed by protein sequence analysis and 'classical' cloning resulted in the identification of some of the enzymes responsible for the activity observed in the whole-cells. However, the sequences of genomes now inform us that multiple enzymes catalyzing the relevant reaction may exist in the same organism and might theoretically contribute to the behaviour of the behaviour of the wild-type whole-cell biocatalyst. The intragenomic complement of that enzyme class within the organism can now in many cases be readily identified after genome sequencing through, for example, sequence homology-based searching such as BLAST [16], often making use of enzyme-specific amino-acid sequence motifs which have roles in catalytic function. Cloning, expression and functional assay of the individual members of the intragenomic complement can then provide information of the contribution of that enzyme to the whole-cell activity.

In many cases the number of enzymes capable of catalyzing an equivalent chemical reaction in an organism can be directly related to its natural biochemical context. Plant genomes, for example, have been shown to contain a large amount of homologues of glycosidases and glycosyltransferases, essential for the breakdown and synthesis of sugars that form part of the cell wall [17]. Microorganisms such as *Streptomyces* [3,11–13], which are active in the biosynthesis of secondary metabolites, are rich in sequences encoding polyketide synthases (PKS), non-ribosomal peptide synthases (NRPS) and enzymes such as cytochromes P450 (CYPs) for the decoration of secondary metabolites resulting from PKS or NRPS activity. Soil organisms of the genus *Pseudomonas* [5] and *Rhodococcus* [8] contain multiple sequences for enzymes that are active in the oxygenation of hydrocarbons, commensurate with their abilities to break down a wide range of structurally diverse carbon sources in order to derive energy for growth. Filamentous fungi such as *Aspergillus* [2,14,15] are rich in hydrolase and oxidase enzymes that assist in the breakdown of biopolymers, consonant with their natural context as saprophytic organisms needing the relevant metabolic armoury to break down cellulose and lignin. In each case the common factor is the evolutionary advantage bestowed on the organism by enzymatic diversity in the face of the wide-ranging chemical challenges that encounter the organism in its natural environment, either in the synthesis or breakdown of natural products.

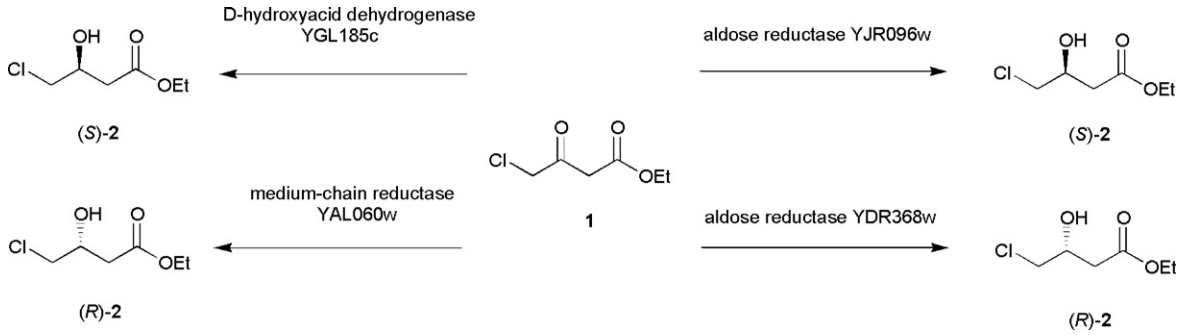
The functional analysis of intragenomic enzyme complements is being driven in part by the desire of the biocatalysis community to unearth new enzymes displaying diverse properties, in order to create a pool on which to draw for application. The summary below focuses primarily on reports where an attempt has been made to characterize, through genome sequence analysis, followed by heterologous expression and functional assay, the catalytic properties of enzyme families within individual genomes. Examples of ketoreductase, hydrolase, glycosyltransferase and oxygenase complements reveal a pleasing range of catalytic diversity within a small number of targets that may constitute a readily accessible library of pre-evolved diversity as a starting point for biocatalyst discovery.

3. Ketoreductases

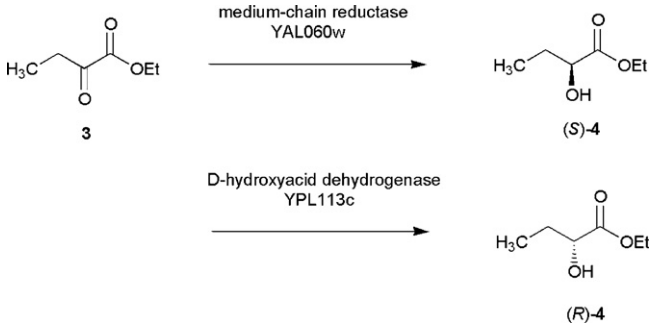
Baker's yeast (*Saccharomyces cerevisiae*) is very well known as a biocatalyst, primarily for its usefulness in the reduction of prochiral ketones to give chiral alcohols [18]. Whilst many successful carbonyl reductions using this and other yeasts continue to be reported, it is still evident that for many of these reactions, the enantioselectivity is compromised by the opposing selectivity of more than one ketoreductase (KRED, alcohol dehydrogenase or carbonyl reductase) enzymes within the wild-type organism. This inherent disadvantage prompted one of the first systematic studies of an intragenomic complement by Stewart and co-workers, who reported the cloning, expression and characterisation of all the open-reading frames encoding KREDs in baker's yeast, in an effort to create 'designer' catalysts that displayed improved selectivity [19]. The genome of *S. cerevisiae*, first published in 1996 [20] appeared to contain forty-nine open reading frames that might be predicted to catalyse carbonyl reductions. Twenty-two were selected that, as a group, included homologs of aldose reductases (AR), short chain reductases (SDR), medium chain reductases (MDR) and D-hydroxyacid dehydrogenases (HAD). The zinc-containing yeast alcohol dehydrogenases were excluded from the study on the basis of being likely to exhibit poor selectivities. In the first instance, eighteen of the twenty-two selected genes were expressed in *E. coli* as GST-fusion tagged proteins and assayed against five representative β-keto ester substrates. It was observed that complementary selectivities were displayed for the reduction of **1** by, for example, YJR096w and YDR368w, each representatives of the aldose reductase superfamily to give (*S*)-**2** and (*R*)-**2** respectively (Scheme 1); Complementary selectivities were also exhibited by enzymes from different superfamilies: YGL185c (D-hydroxyacid dehydrogenase) and YAL060w (medium-chain reductase) again giving (*S*)-**2** and (*R*)-**2** respectively (Scheme 1). Another spectrum of activities was described for a short series of α-keto esters including **3** (Scheme 2), for which YAL060w (MDR) gave (*S*)-**4** and YPL113c (D-hydroxyacid dehydrogenase) gave (*R*)-**4**.

In addition to confirming that the moderate enantioselectivity of whole-cell reductions catalysed by yeast was certainly due to the competing activities of diverse reductases, the study revealed a valuable family of enzymes of complementary selectivities for further application. However, not all the reactions described for whole-cell reactions could be attributed to the activity of these eighteen enzymes. For example, **5** was reduced to (*S*)-**6** by whole cells (Scheme 3), yet none of the GST-fusions of the individual KREDs was able to catalyse that reduction with equivalent selectivity. The study also demonstrated that the relationship between amino acid sequence of the KREDs studied and both the substrate specificity and enantioselectivity was complex, suggesting questionable predictive value in the data for genome mining for catalysts of a particular selectivity in this instance.

A range of substrate spectrum and enantioselectivity was also observed amongst the ketoreductase domains of polyketide synthase modules from *Saccharopolyspora erythrea*, such as



Scheme 1. Enantiocomplementary reductions of a β-keto ester **1** by ketoreductases from *Saccharomyces cerevisiae*.



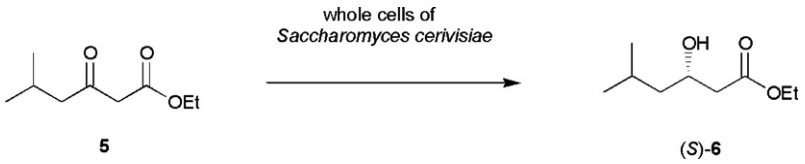
Scheme 2. Enantiocomplementary reductions of a α-keto ester **3** by ketoreductases from *Saccharomyces cerevisiae*.

that which catalyses the assembly of the erythromycin precursor 6-deoxyerythronolide B (DEB synthase or DEBS) [21]. KREDs from modules 1, 2, 5 and 6 were cloned and expressed as GST-fusion proteins, and their extended substrate specificity towards linear aliphatic ketones and cycloalkanones such as cyclohexanone **7** was assessed (Scheme 4) [22]. Each displayed activity towards **7**, racemic 2-allyl cyclohexanone **8**, dicyclohexyl ketone **9** and *trans*-decalone **10**. The KRED from module 1 displayed exceptionally broad substrate specificity, encompassing cyclohexane-1,2-dione **11**, 2-methylcyclohexanone **12**, 2-octanone **13** and 2-methyl-3-heptanone **14**, although activity was lower toward the straight-chain substrates **13** and **14**.

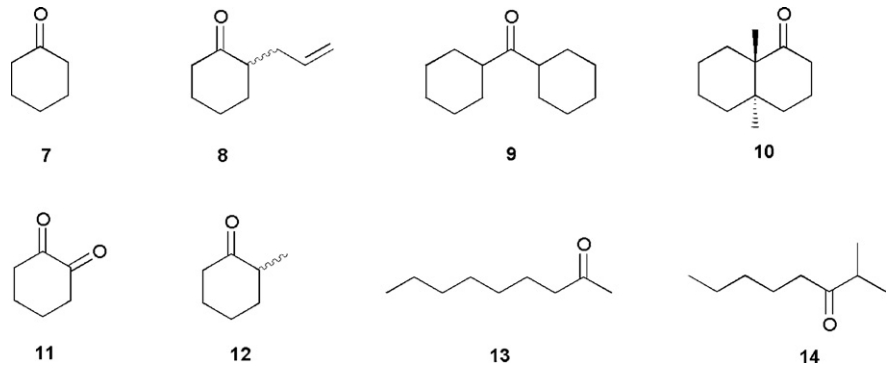
The potential application of such KREDs in biocatalysis would strongly depend on their enantioselective properties. The enantioselectivity of those KRED domains was investigated in detail by Cane and co-workers [23]. It had already been demonstrated [24] that the KRED domains of modules 1, 2, 5 and 6 each transferred the 4-*si*-(4-pro-*S*) hydride of NADPH to the carbonyl group and also that three of those modules, 2, 5 and 6, were responsible for generating L-3-hydroxyacyl-ACP intermediates. In addition, it had been shown that an active site triad of serine, tyrosine and lysine residues was conserved amongst these SDR family members, and also that enzymes displaying D-selectivity possessed an aspartate as part of an LDD motif that was absent from the L-selective members including the KRED domain of DEBS module 4 [25]. Using the *N*-acetyl cysteamide (SNAC) analogues **15** of their natural substrates, the module 1 KRED was confirmed to be L-specific (Scheme 5), but similar studies with KREDs from modules 2, 5, and 6 gave 2–20% of the D-form of the product **16**, suggesting that the known absolute stereoselectivity was due in part to the anchoring of the natural substrate to the acyl carrier protein domain in DEBS.

4. Hydrolases

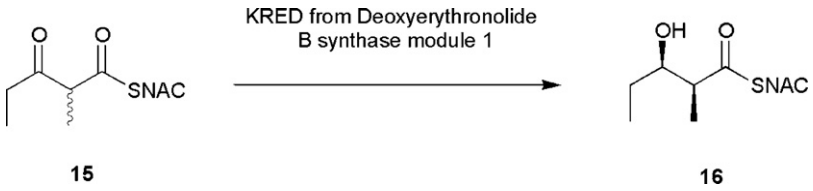
Hydrolytic enzymes, most notably lipases and esterases, form the most important class of enzymes for biocatalytic applica-



Scheme 3. Asymmetric reduction of a β-keto ester **5** by whole-cells of *Saccharomyces cerevisiae*. No isolated enzyme from that organism was described that possessed the same selectivity.



Scheme 4. Compounds tested as substrates for KREDs from deoxyerythronolide B synthase (DEBS) modules from *Saccharopolyspora erythrea*.



Scheme 5. Reduction of *N*-acetyl cysteamide derivative **15** by the module 1 ketoreductase (KRED) from *Saccharopolyspora erythrea*.

tion and it is no surprise that single species contain multiple relevant sequences within their genome. The genome of *Bacillus subtilis*, for example, has been reported to contain eight genes encoding lipases or esterases [26]. Tools for identifying both lipase and esterase enzymes based on sequence motifs have been established [27,28] and have led to the cloning and functional characterization of multiple targets from the same organism, however, some attempts to characterize the whole intragenomic complement of, for example, esterase targets have fallen down because of variable expression and activity. For example, Chung and co-workers identified, cloned and expressed six carboxyl esterases from *Salmonella typhimurium*, yet only two showed significant activity with the standard tributyrin test and only one yielded to detailed analysis [29]. More recently Dherbécourt and co-workers cloned and expressed twelve esterases from *Propionibacterium freudenreichii*, six of which hydrolysed 1-naphthyl acetate, although further exploration of their catalytic diversity was not presented [30]. However, Soror has recently described the cloning and soluble expression of three lipase/esterases from *Streptomyces coelicolor* [31]. SCO7131 was active towards acetate esters (as measured by UV assay), but not larger substrates, was moderately thermostable, and yet non-enantioselective. SCO6966 was active against short-chain fatty acid esters SCO3644 and was not thermostable, indicating that a range of thermostability can be encountered within enzyme families within one genome.

Multiple epoxide hydrolase (EH) genes are also encountered within single genomes. Some of these interesting complements, such as the eight EHs from *Mycobacterium tuberculosis*, are reported to be recalcitrant to expression as a whole [32] although structures of two of those enzymes have been reported [33,34]. However, the expression and functional analysis by GC assay of three EHs from *Erythrobacter litoralis* HTCC2594 hints at the diversity extant within EH families within a single genome [35]. When racemic styrene oxide was used as the substrate, enzyme EEH1 preferentially hydrolysed the (*R*)-enantiomer EEH2 was not enantioselective and EEH3 preferentially hydrolysed the (*S*)-enantiomer. Differences in pH (6.5, 7.5 and 8.0 respectively) and temperature (50 °C, 55 °C and 45 °C) optima were also reflective of a spectrum of catalytic properties within the EH complement of this organism overall.

Zhu and co-workers cloned, expressed and assayed using UV spectrophotometry, 22 of the 23 family GH92 α-mannosidases from *Bacteroides thetaiotamicron* [36], a human gut symbiont which uses a vast array of carbohydrate active enzymes to secure sugars from dietary polysaccharides. The diversity of both substrate specificity and regio- and stereospecificity was notable in this example, with different targets exhibiting α-1,2-mannosidase activity (Bt2199, Bt3962 and Bt3990), or hydrolytic activity towards α-1,3-mannobiose (Bt1769, Bt3858 and Bt3991) or α-1,4-mannobiose (Bt3965, Bt4073 and Bt4093). Catalytic behaviour towards high-mannose N-glycans was also diverse and revealed an array of complex structural requirements of the substrates for recognition by different enzymes. In this case, the diversity observed is again reflective of the catalytic capacity required by the organism to make the most of its available carbon resources for survival.

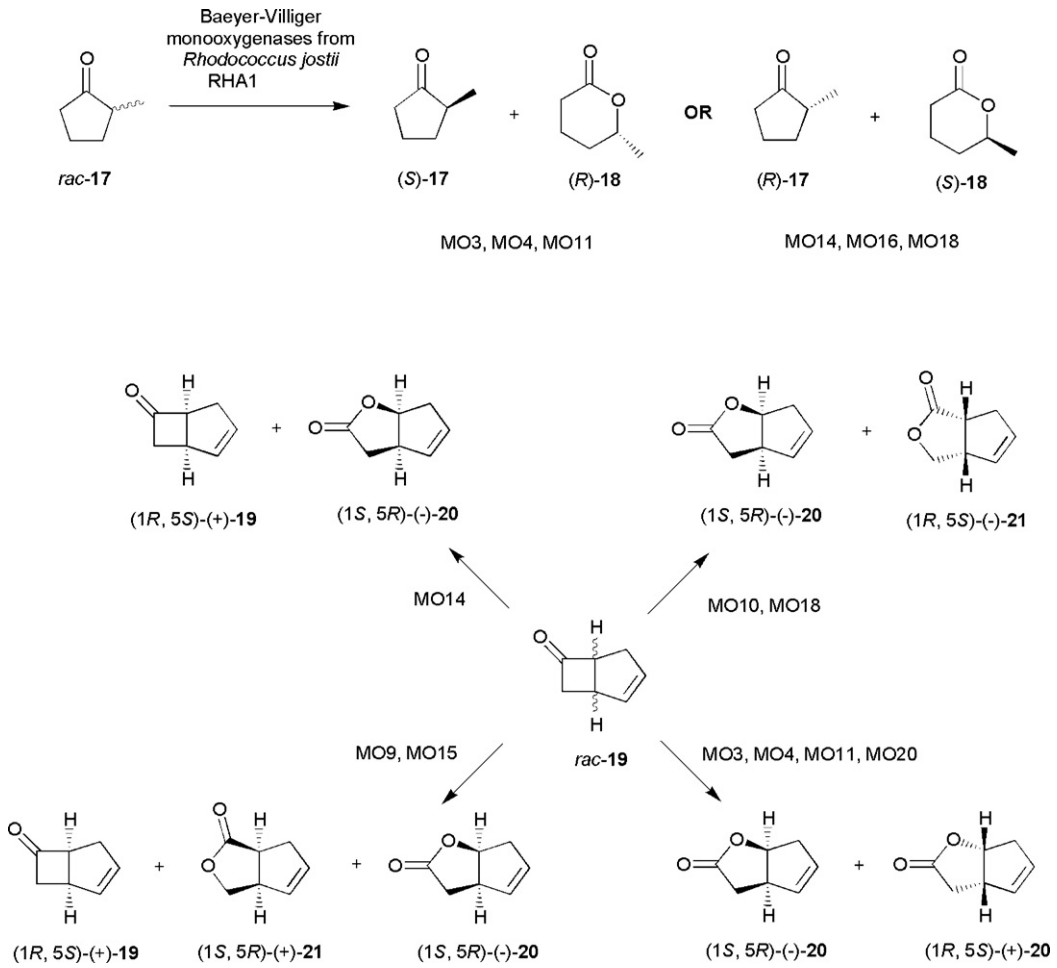
5. Baeyer–Villiger monooxygenases

Baeyer–Villiger monooxygenases (BVMOs) are flavin-containing enzymes that catalyze the oxygenation of ketones to esters or lactones, using reduced nicotinamide coenzyme NADPH to furnish reduced flavin, which reacts with molecular oxygen to form a flavin hydroperoxidate anion that serves as the oxidant in a Baeyer–Villiger reaction [37]. An amino acid sequence motif was identified in 2002 by Janssen and co-workers [38] for identifying Baeyer–Villiger enzymes, and this was successful in identifying, for example, a thermostable BVMO in the genome of *Thermobifida fusca* [39]. It has also proved useful in identifying intragenomic BVMO complements in nocardioform actinomycetes such as *Mycobacterium tuberculosis* [40,41], which contains at least seven BVMO genes, and *Rhodococcus jostii* [8,42].

The genome sequence of the bacterium *Rhodococcus jostii* RHA1 was revealed to encode at least nineteen open-reading frames encoding enzymes with the BVMO motif [8]. Szolkowy and co-workers cloned these, and an additional four related targets from *Rhodococcus* and twelve of them were successfully expressed in the soluble fraction [42]. A screen of substrate acceptance using a GC assay showed that the enzymes accepted a wide range of aliphatic, alicyclic and bicyclic ketones. For the most part, and in contrast to the situation with KREDs from *S. cerevisiae*, amino acid sequence correlated strongly not only with substrate specificity, but also enantioselectivity, for which a spectrum was observed amongst the thirteen targets screened (Scheme 6). For example, groups of BVMOs displayed either (*R*)- or (*S*)-selectivity for the resolution of 2-methylcyclopentanone **17**, mostly with poor *E* values, but one of the targets, termed MO18, catalysed the (*S*)-selective resolution with an *E* value of 25. Bicyclic substrate **19** (Scheme 6) was converted to a range of lactone products by the different enzymes and one of these, termed MO14, was able to catalyse a resolution of the substrate, to give optically pure residual ketone substrate and lactone product. This study also revealed novel families of BVMOs, including those more divergently related to those enzyme previously studied, which featured significant variations in the established amino acid sequence motif.

6. Glycosyltransferases

It has been well documented that, given the fundamental importance of carbohydrate metabolism in biology, that as much as 2% of an organism's genome can encode enzymes involved in either carbohydrate synthesis (glycosyltransferases) or breakdown (glycosidases) [17]. The genome of the plant *Arabidopsis thaliana*, for example, contains at least 356 glycosyltransferases from 27 families [43]. Developments in high-throughput screening of enzyme activity have allowed the synthetic potential of some of this reservoir of enzymes to be investigated. Bowles and co-workers identified a family of 99 genes (including 11 pseudogenes) encoding UDP-glycosyl transferases in *Arabidopsis thaliana* using a C-terminal consensus sequence as an identifier, and with sequence similarities varying between 30 and 95% in 12 major groups [44]. Functions, based on similarity to known enzymes, were proposed which included the glucosylation of anthocyanidins and the plant



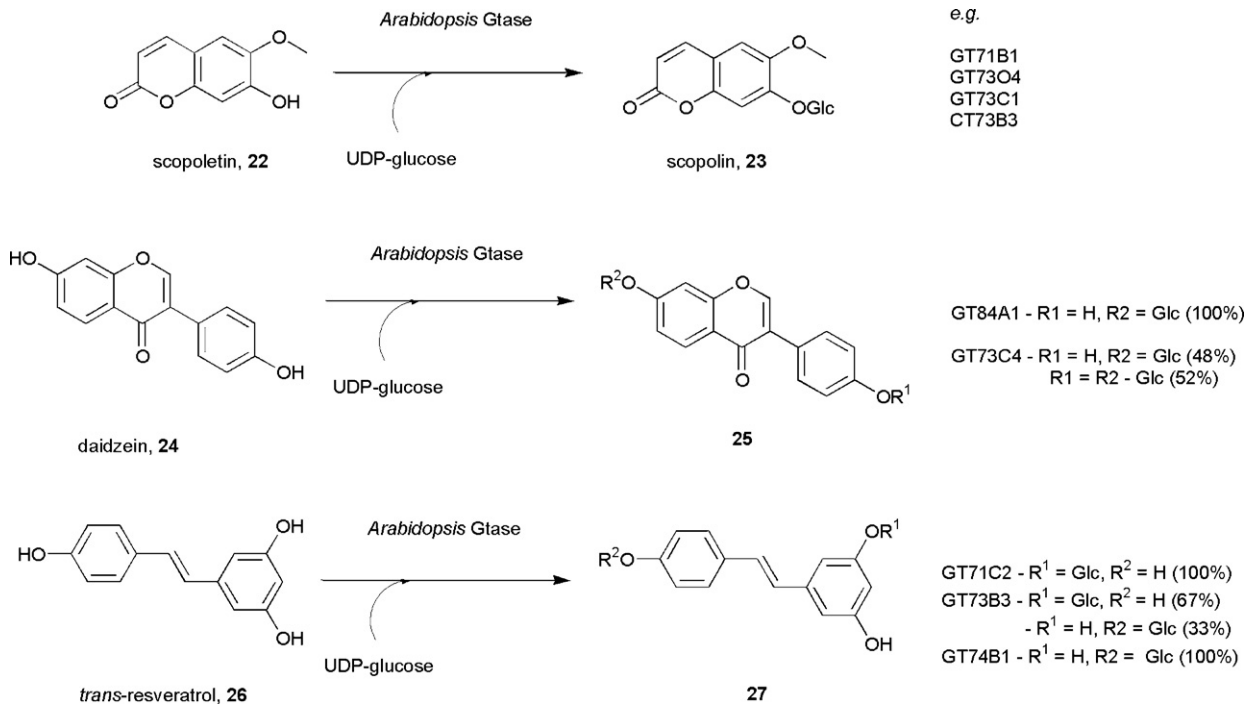
Scheme 6. Spectrum of enantiospecificities displayed by Baeyer–Villiger monooxygenases from *Rhodococcus jostii* RHA1 in the transformation of racemic 2-methylcyclopentanone **17**, and racemic bicyclic ketone **19**.

hormone zeatin. The biocatalytic potential of these enzymes was realized by expressing them in soluble form in *E. coli* [45], leading to a library of 96 active biocatalysts that could be screened in 96-well plate format for the reaction of interest; in that case, the regioselective glucosylation of the flavonoids scopoletin **22**, diadzein **23** and the stilbene derivative *trans*-resveratrol **26** (Scheme 7). In the method described, mixtures of the glycoside products **23**, **25**, and **27** formed were subjected to enzymatic hydrolysis and the resultant glucose detected by a coupled enzyme assay involving glucose oxidase, horseradish peroxidase and a peroxide sensitive dye. Using this method, the library screening revealed that forty-five Gtases from *A. thaliana* demonstrated the ability to glycosylate scopoletin **22**, and twenty-five were competent for the glycosylation of *trans*-resveratrol **26**. Complementary regioselectivities of glycosylation, and also di-glycosylation were observed amongst this intragenomic complement, and those selectivities verified by HPLC analysis. The genome of *A. thaliana* has also allowed the identification of complementary transferases that deliver rhamnose to aglycones such as quercetin [46].

7. Cytochromes P450

Cytochromes-P450 (CYPs) are heme-containing monooxygenases catalyzing a range of oxidation reactions with enormous potential for biocatalytic processes [47]. The 'CYPomes' (intragenomic complement of CYPs) of organisms have therefore already proved to be a popular target for accessing catalytic diversity, with groups having studied the catalytic competence of the relevant complements from bacteria, fungi and also humans. However, in contrast to the other enzyme classes of interest to preparative biocatalysis described above, CYPs are remarkably divergent in reaction chemistry, catalyzing reactions as diverse as carbon hydroxylation, epoxidation, heteroatom dealkylation and oxidation, phenolic couplings and reductive denitrations [48,49]. This catalytic heterogeneity presents challenges when screening CYPomes for reactions of interest, as no 'generic' substrate for their activity has been described. A further challenge is presented by the dependence of CYPs on auxiliary electron transport proteins for activity. For these reasons, whilst the identification of open reading frames encoding CYPs within genomes is relatively simple, based on sequence homology, there is little information on the function of many of the associated gene products. Some studies of CYPomes, such as those of *Mycobacterium tuberculosis* [50] and *Streptomyces coelicolor* [51] have been stimulated by interests in the relevant organism as pathogens or as antibiotic producers respectively, but we focus largely on those CYPomes that have been investigated from the biocatalysis perspective.

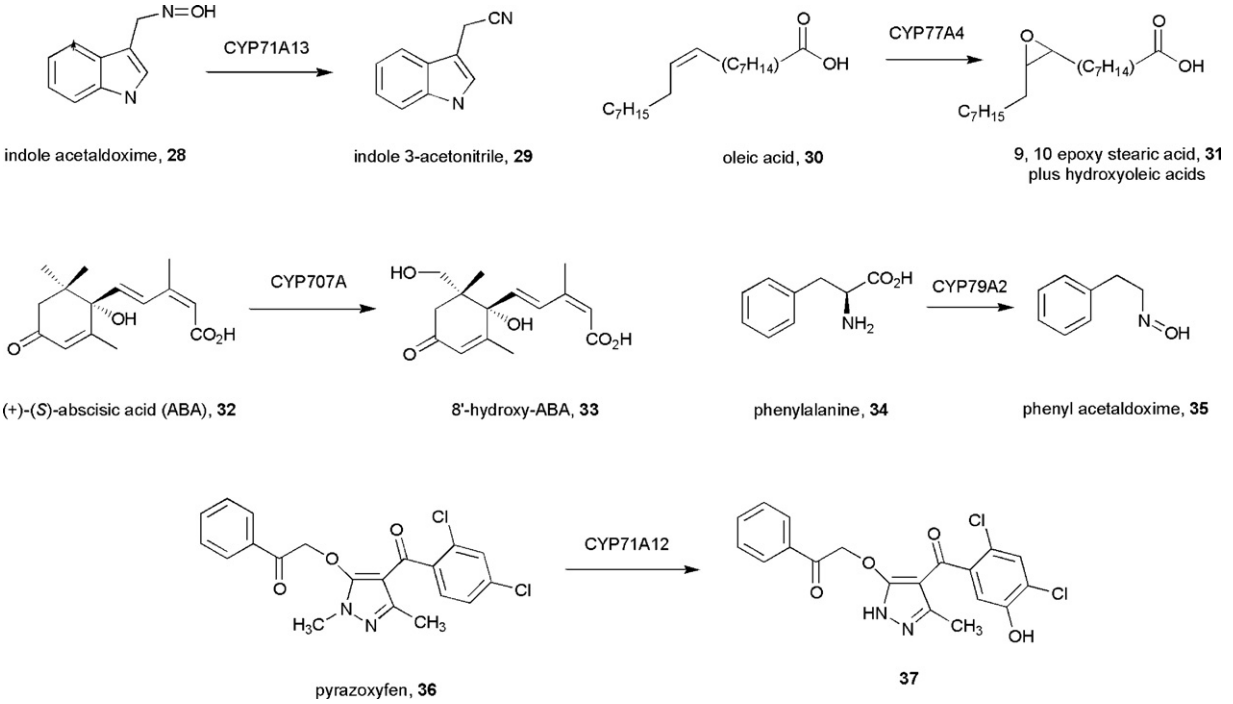
The common soil and water bacterium *Rhodopseudomonas palustris* [52] has seven CYPs encoded within its genome, which are very sequence diverse, representing seven different families. Bell and co-workers cloned and expressed in soluble form in *E. coli* four of these (CYP195A2, CYP199A2, CYP203A1 and CYP153A5) in addition to a ferredoxin associated with CYP199A2, and assessed their ligand binding behaviour with an assay that



Scheme 7. Glucosylation activity of a library of glucosyltransferases from *Arabidopsis thaliana* against aromatic aglycones.

monitors the spectroscopic shift from 418 nm to 390 nm associated with substrate binding at the heme [53]. Whilst no substrates were observed to bind to CYP153A5, CYP199A2 displayed the characteristic shift with aromatic aldehydes and carboxylic acids such as 4-ethylbenzoic acid and 4-isopropylbenzaldehyde. No aliphatic acids were accepted by either CYP199A2 or CYP203A1. An *in vitro* oxidation system of CYP199A2, a ferredoxin from *R. palustris* termed palustrisredoxin, and putidaredoxin reductase was successfully reconstituted for the oxidation of 4-ethylbenzoic acid.

An impressive example of a comprehensive study of the expression of CYPome complements from bacteria is provided by Agematu and co-workers, who cloned and expressed the entire CYP complements of *S. coelicolor* A3(2), *Bacillus subtilis* 168, *Nostoc* sp. PCC7120 and *Nocardia farcinica* IFM 10152 in an attempt to unearth enzymes competent for the selective hydroxylation of testosterone [54]. Four targets from the *Nostoc* CYPome were shown to display contrasting selectivity, with CYPs *alr4686*, *alr4766* and *alr4833* catalysing regioselective hydroxylations of the substrate in the 17, 15β- and



Scheme 8. Reactions confirmed to be catalysed by CYPs from *Arabidopsis thaliana*.

Table 1
Summary of intragenomic enzyme complements catalysing different reactions.

Organism	Enzyme type	No. of targets in genome	No. targets expressed in soluble form	No. of targets active	Notes	Reference
S. cerevisiae	Ketoreductase (SDR)	49	18 (of 22 cloned)	18	(*S*) and (*R*)-selectivity observed	[19]
Propionibacterium freudenreichii	Esterase	23	11 (of 12 cloned)	6	–	[30]
S. coelicolor A3(2)	Lipase/Esterase	31	3 (of 11 cloned)	3	Different thermostabilities observed	[31]
Erythrobacter litoralis	Epoxide Hydrolase	3	3	3	Both (*R*) and (*S*) selectivity observed for styrene oxide hydrolysis	[35]
Bacteroides thetaiotamicron	α-Mannosidase	23	22	22	Different regiospecificities for bond cleavage	[36]
R. jostii	Baeyer–Villiger monooxygenase	23	12	12	Different regioselectivities and enantioselectivities observed	[42]
Arabidopsis thaliana	UDP-Glycosyltransferase	99	96	96	Different regioselectivities towards flavonoid glycosylation	[45]
R. palustris	Cytochrome P450	7	7	4	–	[53]
Nostoc PCC 7120	Cytochrome P450	7	Data not reported	4	Four CYPs with different regioselectivties for testosterone hydroxylation	[54]
Nocardia farcinica	Cytochrome P450	23	12 (cloned and expressed as fusion proteins)	3	–	[55]

6β- positions respectively. Sabbadin and co-workers have further explored the activities of the CYPome from *Nocardia* using a generic fusion enzyme strategy, and uncovered complementary substrate specificities amongst the complement, including enzymes that oxidize testosterone and ethoxycoumarin [55].

The CYPomes of eukaryotes present additional challenges in respect of functional expression and characterization, being membrane bound in their natural state, and dependent on the activity of CYP reductase proteins that are responsible for transferring reducing equivalents from NADPH to the heme domain. Filamentous fungi, widely used for industrial hydroxylation reactions, are proving to be rich sources of genes encoding CYPs [56]. Whilst there are a few reports of the characterization of fungal CYPomes *in silico* including that of *Aspergillus nidulans* [57] attempts at the functional characterisation of fungal CYPomes are limited. Wariishi and co-workers recently cloned 121 CYPs from *Aspergillus oryzae* although functional characterization of the enzymes is outstanding [58]. More work has been accomplished on plant CYPomes, however, which constitute a compelling array of potential activities for biocatalysis when one considers the multiple roles for CYPs in the biosynthesis of, for example, bioactive alkaloids and flavonoids, The CYPome of *Arabidopisis thaliana* has been the most extensively studied, and contains at least 272 CYPs containing members of over 45 families (http://www.p450.kvl.dk/Arab_cyps/family.shtml). Whilst a comprehensive description of the activities of *Arabidopsis* CYPs is frustrated by the practical problems associated with the expression and assay of these enzymes, some clues in respect of functional annotation have been provided by large-scale co-expression analysis [59] using a large set of publicly available databases. This analysis allowed Werck-Reichhart and co-workers to suggest roles for some of the *Arabidopsis* CYPs in some metabolic pathways, notably the biosynthesis of steroids, isoprenoids and triterpenoids. The biochemical role of some individual *Arabidopsis* CYPs has been unambiguously demonstrated using individual cloning and functional characterisation experiments. CYP71A13 catalyses the oxidation of indoleacetaldoxime **28** to form indole-3-acetonitrile **29** (Scheme 8) in the biosynthetic pathway towards the alkaloid camalexin [60]; CYP77A4 catalyses the epoxidation of fatty acids such as oleic acid **30** and linoleic acid [61]; CYP707A catalyses the 8'-hydroxylation of abscisic acid **32** [62]. CYP79A2 catalyses the transformation of phenylalanine **34** to phenylacetaldoxime **35** as part of the pathway towards benzylglucosinate [63]. Other *Arabidopsis* CYPs have been shown to have activity towards xenobiotics, such as CYP71A12 which metabolises the herbicide pyrazoxyfen **36** [64]. A genomics project designed to collate information on the functional assignment of *Arabidopsis* CYPs can be found at http://arabidopsis-p450.biotec.uiuc.edu/, including a regularly updated list of functions which have been unambiguously defined. It is already clear that such a vast range of sequences encodes a suitably diverse spectrum of CYP activity for *Arabidopsis* to fulfil its metabolic and biosynthetic roles.

8. Conclusion – smart libraries for biocatalyst discovery?

The data emerging from the functional expression of families of enzymes catalyzing equivalent chemical reactions in single genomes are beginning to confirm the existence of a wide range of catalytic properties that might be expected as a result of evolutionary adaptation of each organism to the demands of its environment. In many examples it is the drive to discover new biocatalysts that is helping to reveal this new information on the catalytic capacity of organisms, and is throwing up some examples of interesting new selectivities along the way. The intragenomic complements studied thus far (some of which are summarized in Table 1) might be considered subsets of activities that can be targeted as discrete pools of 'pre-evolved' sequences from which to begin a search for enzymes of new and useful activities for novel application. For example, in the case of *Saccharomyces* KREDs, *Erythrobacter* EHs and *Rhodococcus* BVMOs, these small complements have certainly revealed enzymes of contrasting enantioselectivities, each of which may serve as a valuable starting point for *in vitro* enzyme evolution for catalytic optimization and process suitability. There are of course be limitations to an enzyme discovery approach based only on an intragenomic complement: The enzyme family must have several members within the genome, and one must be fairly confident that each member has at least the potential to catalyse the desired chemical reaction of interest. This can be difficult to predict, as illustrated by the case of low-sequence ID superfamilies such as the CYPs in *Arabidopsis*. Also, the expression or activity of the homologues, once cloned can prove to be variable as with all gene targets, as shown with the esterases of *S. coelicolor* or the CYPs of *Nocardia* (Table 1). Certainly a choice of twenty sequence-diverse targets from one organism may not be superior to twenty targets from twenty different organisms, or to the diversity accessible through metagenomics approaches. However, an intragenomic complement of enzymes is attractive, particularly from the academic research perspective, in employing a single DNA sample required for amplification of the relevant targets, and, possibly generic PCR and expression protocols, allowing rapid access

to a starting pool of diversity and both practical and economical advantages. In the future, such advantages might be made redundant by inexpensive gene synthesis technology, but it is interesting nevertheless at this time to observe how the discipline of applied biocatalysis and the search for new enzymes is helping to shed light on the power and diversity of enzymatic catalysis within individual species.

References

[1] J. Kalinowski, B. Bathe, A. Tauch, et al., J. Biotechnol. 104 (2003) 5–25.
[2] H.J. Pel, J.H. de Winde, H. Stam, et al., Nat. Biotechnol. 25 (2007) 221–231.
[3] S. Omura, H. Ikeda, M. Hattori, et al., Proc. Natl. Acad. Sci. 98 (2001) 12215–12220.
[4] M. Ferrer, P.N. F-Martínez-Abarca, Golyshin, Curr. Opin. Biotechnol. 16 (2005) 588–593.
[5] K.E. Nelson, C. Weinel, C.M. Fraser, et al., Environ. Microbiol. 12 (2002) 799–808.
[6] L.P. Wackett, Nat. Biotechnol. 21 (2003) 136–138.
[7] F. Kunst, N. Ogasawara, A. Danchin, et al., Nature 390 (1997) 249–256.
[8] M.P. McLeod, R.L. Warren, L.D. Eltis, et al., Proc. Natl. Acad. Sci. 103 (2006) 15582–15587.
[9] J. Ishikawa, A. Yamashita, Y. Mikami, Y. Hoshino, H. Kurita, K. Hotta, T. Shiba, M. Hattori, Proc. Natl. Acad. Sci. 101 (2004) 14925–14930.
[10] E.F. Mongodin, N. Shapir, S.C. Daugherty, R.T. DeBoy, J.B. Emerson, A. Shvartzbeyn, D. Radune, J. Vamathevan, F. Riggs, V. Grinberg, H. Khouri, L.P. Wackett, K.E. Nelson, M.J. Sadowsky, PloS Genet. 2 (2006) e214.
[11] S.D. Bentley, K.F. Chater, D.A. Hopwood, et al., Nature 417 (2002) 141–147.
[12] Y. Ohnishi, J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A. Yamashita, M. Hattori, S. Horinouchi, J. Bacteriol. 190 (2008) 4050–4060.
[13] R. Kirby, T.-K. Gan, I. Hunter, P. Herron, E. Tilley, Antonie van Leeuwenhoek 94 (2008) 173–186.
[14] J.E. Galagan, S.E. Calvo, B.W. Birren, et al., Nature 438 (2005) 1105–1115.
[15] M. Machida, K. Asai, H. Kikuchi, et al., Nature 438 (2005) 1157–1161.
[16] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (2007) 3389–3402.
[17] P.M. Coutinho, M. Stam., E. Blanc, B. Henrissat, Trends Plant Sci. 8 (2003) 563–565.
[18] J.D. Stewart, Curr. Opin. Drug. Discovery Development 1 (2008) 278–289.
[19] I.A. Kaluzna, T. Matsuda, A.K. Sewell, J.D. Stewart, J. Am. Chem. Soc. 126 (2004) 12827–12832.
[20] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, Science 274 (1996) 546–567.
[21] M. Oliynyk, M. Samborskyy, J.B. Lester, T. Mironenko, N. Scott, S. Dickens, S.F. Haydock, P.F. Leadlay, Nature Biotechnol. 25 (2007) 447–453.
[22] S. Bali, H.M. O'Hare, K.J. Weissman, Chembiochem 7 (2006) 478–484.
[23] R. Castonguay, W. He, A.Y. Chen., C. Khosla, D.E. Cane, J. Am. Chem. Soc. 129 (2007) 13758–13769.
[24] M. McPherson, C. Khosla, D.E. Cane, J. Am. Chem. Soc. 120 (1998) 3267–3268.
[25] P. Caffrey, Chembiochem 4 (2003) 649–662.
[26] E. Henke, U.T. Bornscheuer, Appl. Microbiol. Biotechnol. 60 (2002) 320–326.
[27] H.K. Kim, Y.-J. Jung, W.-C. Choi, H.S. Ryu, T.-K. Oh, J.-K. Lee, FEMS Microbiol. Lett. 235 (2004) 349–355.
[28] H.-Y. Kang, J.F. Kim, M.H. Kim, S.-H. Park, T.-K. Oh, C.-G. Hur, FEBS Lett. 580 (2006) 2736–2740.
[29] H.-S. Ro, H.P. Hong, B.H. Kho, S. Kim, B.H. Chung, FEMS Microbiol. Lett. 233 (2004) 97–105.
[30] J. Dherbécourt, H. Falentin, S. Canaan, A. Thierry, Microb. Cell. Fact 7 (2008) 16.
[31] S. Soror, Ph.D. thesis, University of Kaiserslauten, 2007.
[32] B. van Loo, J. Kingma, M. Arand, M.G. Wubbolts, D.B. Janssen, Appl. Environ. Microbiol. 72 (2006) 2905–2917.
[33] P. Johansson, T. Unge, A. Cronin, M. Arand, T. Bergfors, T.A. Jones, S.L. Mowbray, J. Mol. Biol. 351 (2005) 1048–1056.
[34] B.K. Biswal, C. Morisseau, G. Garen, M.M. Cherney, C. Garen, C. Niu, B.D. Hammock, M.N. James, J. Mol. Biol. 381 (2008) 897–912.
[35] J.-H. Woo, Y.-O. Hwang, S.G. Kang, H.S. Lee, J.-C. Cho, S.-J. Kim, Appl. Microbiol. Biotechnol. 76 (2007) 365–375.
[36] Y. Zhu, M.D.L. Suits, A.J. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K.W. Moremen, Y. Xiang, A. Siriwardena, S.J. Williams, H.J. Gilbert, G.J. Davies, Nat. Chem. Biol. 6 (2009) 125–132.
[37] M.D. Mihovilovic, Curr. Org. Chem. 10 (2006) 1265–1287.
[38] M.W. Fraaije, N.M. Kamerbeek, W.J.H. van Berkel, D.B. Janssen, FEBS Lett. 518 (2002) 43–47.
[39] M.W. Fraaije, J. Wu, D.P.H.M. Heuts, E.W. van Hellemond, J.H. Lutje Spelberg, D.B. Janssen, Appl. Microbiol. Biotechnol. 66 (2004) 393–400.
[40] M.W. Fraaije, N.M. Kamerbeek, A.J. Heidekamp, R. Fortin, D.B. Janssen, J. Biol. Chem. 279 (2004) 3354–3360.
[41] D. Bonsor, S.F. Butz, J. Solomons, S.C. Grant, I.J.S. Fairlamb, M.J. Fogg, G. Grogan, Org. Biomol. Chem. 4 (2006) 1252–1260.
[42] C. Szolkowy, L.D. Eltis, N.C. Bruce, G. Grogan, Chembiochem 10 (2009) 1208–1217.
[43] B. Henrissat, P.M. Coutinho, G.J. Davies, Plant. Mol. Biol. 47 (2001) 55–72.
[44] Y. Li, S. Baldauf, E.-K. Lim, D.J. Bowles, J. Biol. Chem. 276 (2001) 4338–4343.
[45] M. Weis, E.-K. Lim, N. Bruce, D. Bowles, Angew. Chem. Int. Ed. 118 (2006) 3614–3618.
[46] E.K. Lim, D.A. Ashford, D.J. Bowles, Chembiochem 7 (2006) 1181–1185.
[47] R. Bernhardt, J. Biotechnol. 124 (2006) 128–145.
[48] F.P. Guengerich, Chem. Res. Toxicol. 14 (2001) 611–650.
[49] A.W. Munro, H.M. Girvan, K.J. McLean, Nat. Prod. Rep. 24 (2007) 585-609.
[50] H. Ouellet, J.B. Johnston, P.R. Ortiz de Montellano, Arch. Biochem. Biophys. 493 (2010) 82–95.
[51] D.C. Lamb, T. Skaug, H.-L. Somg, C.J. Jackson, L.M. Podust, M.R. Waterman, D.B. Kell, D.E. Kelly, S.L. Kelly, J. Biol. Chem. 277 (2002) 24000–24005.
[52] F.W. Larimer, P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M.L. Land, D.A. Pelletier, J.T. Beatty, A.S. Lang, F.R. Tabita, J.L. Gibson, T.E. Hanson, C. Bobst, J.L. Torres y Torres, C. Peres, F.H. Harrison, J. Gibson, C.S. Harwood, Nature Biotechnol. 22 (2003) 55–61.
[53] S.G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L.-L. Wong, Biochem. Biophys. Res. Commun. 342 (2006) 191–196.
[54] H. Agematu, N. Matsumoto, Y. Fujii, H. Kabumoto, S. Doi, K. Machida, J. Ishikawa, A. Arisawa, Biosci. Biotechnol. Biochem. 70 (2006) 307–311.
[55] F. Sabbadin, R. Hyde, A. Robin, E.-M. Hilgarth, M. Delenne, S. Flitsch, N. Turner, G. Grogan, N.C. Bruce, Chembiochem 11 (2010) 987–994.
[56] B. Cresnar, S. Petric, Biochim. Biophys. Acta (2010) doi:10.1016/j.bbapap.2010.06.020.
[57] D.E. Kelly, N. Krasevec, J. Mullins, D.R. Nelson, Fungal Genet. Biol. 46 (2009) S53–S61.
[58] K. Nazmul Hussain Nazir, H. Ichinose, H.H. Wariishi, Arch. Microbiol. 192 (2010) 395–408.
[59] J. Ehling, N.J. Provart, D. Werck-Reichhart, Biochem. Soc. Trans. 34 (2006) 1192–1197.
[60] E. Glawschnig, Biochem. Soc. Trans. 34 (2006) 1206–1208.
[61] V. Sauveplane, S. Kandel, P.E. Kastner, J. Ehlting, V. Compagnon, D. Werck-Reichhart, F. Pinot, FEBS.J. 276 (2009) 719–735.
[62] T. Kushiro, M. Okamoto, K. Nakabayashi, K. Yamagishi, S. Kitamura, T. Asami, N. Hirai, T. Koshiba, Y. Kamiya, E. Nambara, EMBO J 23 (2004) 1647–1656.
[63] U. Wittstock, B.A. Halkier, J. Biol. Chem. 275 (2000) 14659–14666.
[64] E. Hayashi, K. Fuzimoto, H. Imaishi, Plant Biotechnol. 24 (2007) 393–396.