Article

Prospecting Biotechnologically-Relevant Monooxygenases from Cold Sediment Metagenomes: An In Silico Approach

Matías A. Musumeci [1], Mariana Lozada [1], Daniela V. Rial [2], Walter P. Mac Cormack [3,4], Janet K. Jansson [5], Sara Sjöling [6], JoLynn Carroll [7,8] and Hebe M. Dionisi [1,*]

[1] Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos, CONICET, Puerto Madryn, Chubut U9120ACD, Argentina; musumeci@cenpat-conicet.gob.ar (M.A.M.); lozada@cenpat-conicet.gob.ar (M.L.)
[2] Área Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, CONICET, Suipacha 531 S2002LRK Rosario, Argentina; drial@fbioyf.unr.edu.ar or rial@inv.rosario-conicet.gov.ar
[3] Instituto Antártico Argentino, Ciudad Autónoma de Buenos Aires C1010AAZ, Argentina; wmac@ffyb.uba.ar
[4] Instituto de Nanobiotecnología (NANOBIOTEC), CONICET—Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires C1113AAD, Argentina
[5] Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA; janet.jansson@pnnl.gov
[6] School of Natural Sciences and Environmental Studies, Södertörn University, 141 89 Huddinge, Sweden; sara.sjoling@sh.se
[7] Akvaplan-niva, Fram—High North Research Centre for Climate and the Environment, NO-9296 Tromsø, Norway; jolynn.carroll@akvaplan.niva.no
[8] ARCEx—Research Centre for Arctic Petroleum Exploration, Department of Geosciences, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
* Correspondence: hdionisi@cenpat-conicet.gob.ar; Tel.: +54-280-4883-184 (ext. 1251); Fax: +54-280-4883-543

Abstract: The goal of this work was to identify sequences encoding monooxygenase biocatalysts with novel features by in silico mining an assembled metagenomic dataset of polar and subpolar marine sediments. The targeted enzyme sequences were Baeyer–Villiger and bacterial cytochrome P450 monooxygenases (CYP153). These enzymes have wide-ranging applications, from the synthesis of steroids, antibiotics, mycotoxins and pheromones to the synthesis of monomers for polymerization and anticancer precursors, due to their extraordinary enantio-, regio-, and chemo- selectivity that are valuable features for organic synthesis. Phylogenetic analyses were used to select the most divergent sequences affiliated to these enzyme families among the 264 putative monooxygenases recovered from the ~14 million protein-coding sequences in the assembled metagenome dataset. Three-dimensional structure modeling and docking analysis suggested features useful in biotechnological applications in five metagenomic sequences, such as wide substrate range, novel substrate specificity or regioselectivity. Further analysis revealed structural features associated with psychrophilic enzymes, such as broader substrate accessibility, larger catalytic pockets or low domain interactions, suggesting that they could be applied in biooxidations at room or low temperatures, saving costs inherent to energy consumption. This work allowed the identification of putative enzyme candidates with promising features from metagenomes, providing a suitable starting point for further developments.

Keywords: bacterial cytochrome P450; Baeyer–Villiger monooxygenases; bioprospecting biocatalysts; phylogenetic analysis; molecular modeling

1. Introduction

The biotechnological potential of marine bacteria has been exploited in several patented technological processes based on marine enzymes [1]. Metagenomic approaches in poorly characterized marine and coastal environments constitute a promising strategy for the discovery of novel biocatalysts [2]. However, bioprospecting efforts can be hindered by low-coverage sequence information and inefficient read assembly of shotgun sequenced metagenomes as a result of the high diversity of microbial communities in these environments, in particular in sediments [3]. These datasets are comprised mostly of unassembled reads and short scaffolds containing partial protein coding sequences (PCS), with limited biotechnological value. The identification of biocatalyst sequences can also be limited by the low coverage of many metagenomic datasets, as several attractive target enzymes are often encoded in low-abundance members of these highly diverse communities [4]. These methodological restrictions limit the exploitation of the remarkable amount of biotechnologically relevant genetic resources from yet-to-be cultured microorganisms that are currently stored in public metagenomic databases. Furthermore, once a set of sequences is retrieved from the dataset, a knowledge-based selection is required before the synthesis, heterologous expression and characterization of the biocatalyst candidates, in order to avoid the screening of a large number of enzymes, which is expensive and time-consuming [5]. The increasing availability of metagenomic data, coupled to improvements in the design and prediction of protein structures will certainly contribute to improving the initialization steps of directed evolution of protein biocatalysts [6].

Oxygenases, which catalyze the addition of oxygen atoms into many organic compounds, show a remarkable enantio- and regio-selectivity and broad substrate specificity [7–10]. These features make them valuable biocatalysts for the production of synthons relevant for pharmaceutical and chemical industries [4], and it has been suggested that the use of these enzymes will be as prominent as the well-established hydrolases and dehydrogenases [7]. Furthermore, their application in chemical synthesis can replace the use of potentially harmful chemicals (i.e., green chemistry) [8]. Among the most promising group of oxygenases are the Baeyer–Villiger Monooxygenases (BVMOs), well known for their biotechnological potential in the development of pharmaceuticals [11], as well as bacterial cytochrome P450 Monooxygenases (CYP153) [12], which are novel P450 and unique enzymes of the family being soluble and able to hydroxylate highly hydrophobic alkanes [13].

Type I Baeyer–Villiger Monooxygenases (BVMOs) belong to the Group B Flavin-dependent Monooxygenases (B-FDM), a diverse family of enzymes that also includes Flavoprotein Monooxygenases (FMO), N-hydroxylating Monooxygenases (NHMO) and YUCCA Monooxygenases (currently defined as indole 3-pyruvate monooxygenase, EC 1.14.13.168) [14]. BVMOs catalyze the enantio-selective oxidation of ketones to produce esters or lactones, and their applications in organic synthesis have significantly expanded over the last two decades, currently reaching the multi-kilogram scale, and further scale-ups and industrial applications may be expected in the near future [7]. BVMOs have been applied in the production of chiral intermediates for the synthesis of natural products and analogs [15–17], the production of Esomeprazole by Codexis being one remarkable example of BVMOs in the industry [18]. With the current trend to evolve the traditional chemical operations toward environmentally harmless processes, the enzymatic Baeyer–Villiger oxidation is promising for a sustainable development of the chemical industry [19].

CYP153 are soluble haem-containing monooxygenases, which catalyze the hydroxylation of medium-chain-length alkanes (from C6 to C11) at the terminal position to produce 1-alkanols. Applications of CYP153 enzymes include the conversion of terpene limonene into perillyl alcohol, a putative anticancer agent as well as the capability to hydroxylate piperidines, pyrrolidines and azetidines to useful pharmaceutical intermediates [13]. Moreover, in the last few years, high sequence diversity was uncovered in genomes and metagenomes for this enzyme group, highlighting their still unexplored potential [20].

The goal of this work was to identify Baeyer–Villiger and CYP153 Monooxygenases in a metagenomic dataset obtained from shotgun metagenomic sequencing of polar and subpolar coastal sediments [21],

and to select sequences presenting promising biotechnological features such as wide substrate range, regio-selectivity or novel specificities. Starting from a set of more than 200 putative monooxygenases and using a multi-step approach including phylogenetic analysis and molecular modeling, we selected a total of five sequences, four BVMOs and one CYP153, with evidence of broad or novel substrate specificity. This knowledge-based approach is applicable to other biocatalysts and environments and can provide a systematic framework for bioprospecting efforts in complex metagenomic datasets.

2. Results and Discussion

2.1. Identification of Metagenomic Sequences

The dataset used in this work for the identification of sequences homologous to monooxygenase enzymes with biotechnological potential contains 23 assembled metagenomes of coastal sediments from four distant polar or subpolar environments (Advent fjord in Spitsbergen, Värtahamnen in the Baltic Sea; Ushuaia Bay in Tierra del Fuego Island and Potter Cove of 25 de Mayo Island). Both marine and brackish ecosystems exposed to oil-pollutants and cold temperatures are represented in this dataset [21]. Microorganisms adapted to extreme conditions are known to produce enzymes with promising features for biotechnological applications [22].

The dataset was queried for sequences encoding putative BVMOs and CYP153 enzymes using Pfam domains and BLAST (Basic Local Alignment Search Tool) searches (Table S1). The number of identified sequences per metagenome varied widely (Table 1). A positive correlation was detected between the total number of oxygenase sequences identified in the metagenomes and the number of PCSs in the assembled metagenomes (Pearson correlation $r = 0.667$, $p = 0.01$). This result suggests that the number of retrieved sequences could have been limited by the relatively low sequence depth (one lane of Illumina HiSeq 1500 per sample), which probably affected the assembly efficiency (Table 1). This is a common problem in metagenomes of samples with high diversity [23]. In spite of this limitation, 264 monooxygenase sequences were identified in the dataset.

Table 1. Number of retrieved monooxygenase sequences in polar and sub-polar coastal sediment metagenomes.

Sample	B-FDM [a]	CYP153 [b]	Identified Sequences/Metagenome	PCS [c]	Assembly (%) [d]
NOR02	0	0	0	2.82×10^5	12.99
NOR05	8	3	11	7.85×10^5	22.74
NOR08	11	15	26	1.28×10^6	29.04
NOR13	2	1	3	3.13×10^5	17.00
NOR15	9	7	16	1.47×10^6	34.55
NOR18	3	1	4	3.92×10^5	23.55
SWE02	2	5	7	1.11×10^6	19.29
SWE07	0	0	0	2.68×10^5	8.77
SWE12	7	1	8	1.08×10^6	18.17
SWE21	6	3	9	8.66×10^5	17.10
SWE26	0	0	0	5.39×10^5	12.03
ARG01	0	1	1	1.39×10^5	4.50
ARG02	0	1	1	1.74×10^5	5.20
ARG03	0	0	0	4.74×10^5	13.86
ARG04	1	3	4	2.79×10^5	8.15
ARG05	6	8	14	7.13×10^5	13.81
ARG06	0	0	0	1.87×10^5	5.21
ANT01	19	32	51	1.11×10^6	29.19
ANT02	10	23	33	9.72×10^5	27.56
ANT03	1	4	5	2.79×10^5	13.90
ANT04	6	19	25	6.57×10^5	26.45
ANT05	6	12	18	4.67×10^5	20.40
ANT06	11	17	28	1.01×10^5	7.80
Total	108	156	264	1.39×10^7	

[a] B-FDM: Flavin-Dependent Monooxygenases; [b] CYP153: Bacterial Cytochrome P450; [c] PCS: Number of protein coding sequences in the assembled metagenomes; [d] Assembly (%) as percentage of reads mapping the scaffolds.

*2.2. BVMOs*

2.2.1. Structural Modeling

Considering that some members of B-FDM different from BVMO share the same Pfam domain (PF00743) or have a high sequence homology against BVMOs sequences, the identified metagenomic sequences were further classified by phylogenetic analysis using reference sequences belonging to the different subgroups of B-FDM (Figure S1). The most abundant subgroup represented in this set of metagenomic sequences was BVMO, with 36 PCSs unequivocally assigned to different well-supported clusters containing BVMO references (Figure S1). A cluster of 29 sequences grouped with FMO reference sequences, and 40 metagenomic sequences could not be reliably assigned to any subgroup.

With the aim of identifying novel features related to active sites or ligand specificities, a subset of divergent putative BVMO sequences was chosen for further analysis, based on the following selection criteria.

First, clusters supported by bootstrap values higher than 50% and containing at least one full-length metagenomic sequence were identified in the phylogenetic tree (clusters 77, 96, 89, 98 and 59, Figure S1 and Figure 1). Second, the presence of conserved domains consisting of two Rossmann motifs (GxGxx[G/A]) flanking two BVMO fingerprints ([A/G]GxWxxxx[F/Y]P[G/M]xxxD and FxGxxxHxxxW[P/D]) [24–26] was verified in order to confirm their classification as BVMO (Figure 1). Minor differences were observed in some sequences and could be attributed to the divergence of the sequences, considering that residues essential for catalysis remained conserved [24,25]. Full-length metagenomic sequences were then selected from each cluster for modeling the three-dimensional protein structure in detail (Figure 1). In clusters containing metagenomic sequences sharing high identity values (>90.00%), only one representative sequence was selected for structural modeling. Structural models were calculated by homology modeling. Template structures were selected considering coverage, homology percentage and resolution (Table S2). For each metagenomic sequence, 15 models were constructed based on the alignments of models generated using MODELLER, PROMALS3D and MAFFT, and manual adjustments were performed when necessary in order to attain alignment accuracy. The model with the best quality parameters was selected for each of the twelve metagenomic sequences. These selected models were overlapped with their respective template structures for structural comparison (Figure 2A). The structural models of the putative BVMOs showed overall folds similar to those of template structures, with differences in some of the loops that connect α-helixes or β-strands, especially in the so-called "Control Loop" [27] (Figure 2A). This loop influences the active site environment and plays a critical role in enzyme structure and catalysis, mediating NADPH (Nicotinamide Adenine Dinucleotide Phosphate, hydride reduced form) binding and substrate selection [27].

Amino acid residues identified as relevant for catalysis, substrate interaction or stereo-selectivity in crystallized BVMOs [28–31] were compared with the corresponding residues in the models of metagenomic putative BVMOs (Figure 2B). The residues relevant for catalysis (D, R and W) were conserved in the analyzed metagenomic structural models. Although tyrosines or phenylalanines were observed in some models instead of a conserved catalytic tryptophan involved in NADPH-interaction (W492, 3GWD numbering), the aromatic nature was preserved. However, this difference could influence the co-factor binding mode. More significant differences were observed in residues involved in substrate interaction or regioselectivity (Figure 2A).

These results suggest that the putative BVMOs identified in the metagenomic dataset could present differences in their substrate profile and/or stereo-selectivity, as well as their NADPH binding mode.
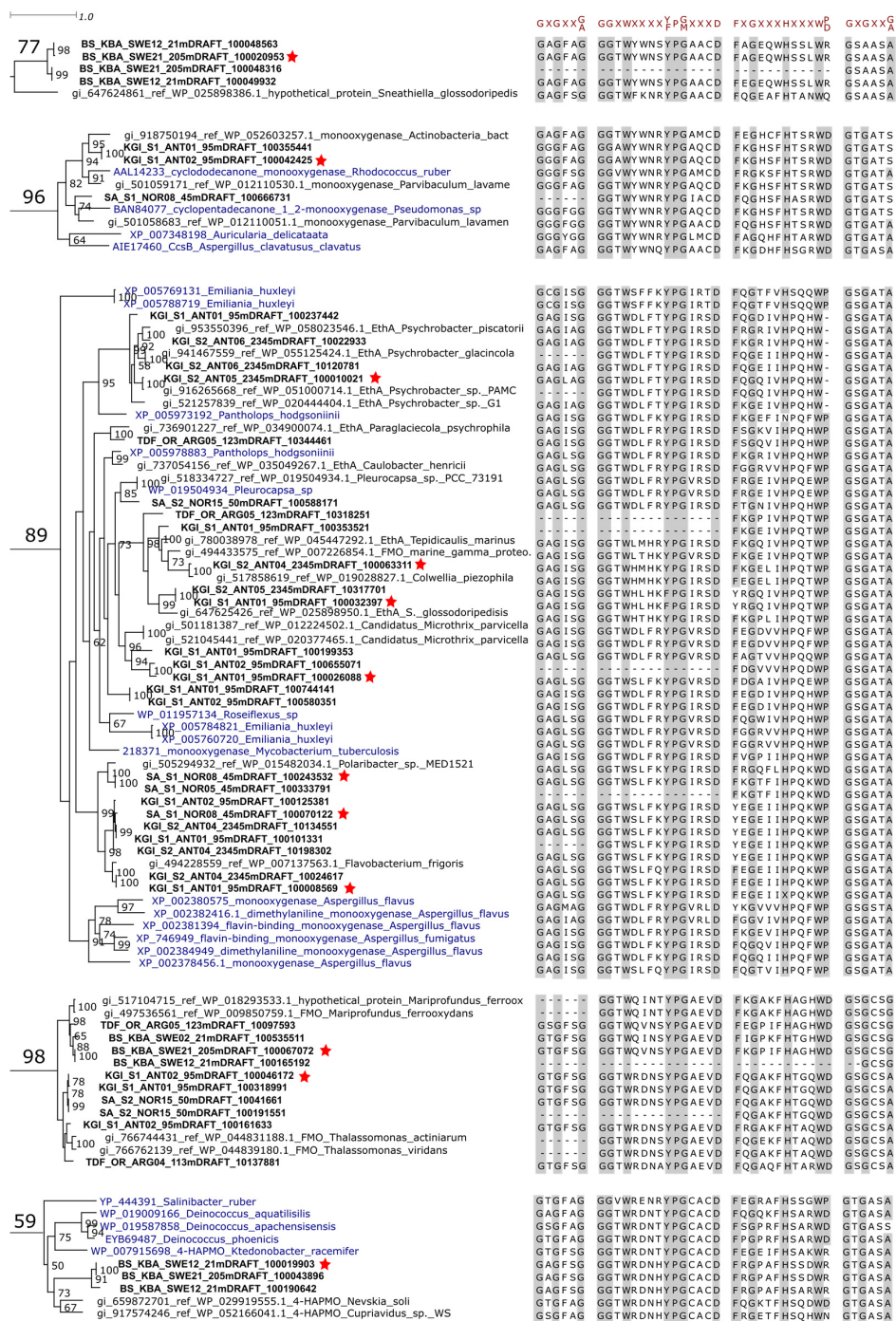
Figure 1. Phylogenetic relationships between metagenomic sequences and BVMO (Baeyer-Villiger Monooxygenase) reference sequences in selected clusters. Clusters were chosen based on the presence of sequences of interest (the full maximum-likelihood phylogenetic tree is shown in Figure S1). Metagenomic sequences and reference BVMO sequences [14,24] are shown in bold and blue, respectively. First matches of metagenomic sequences in BLAST (Basic Local Alignment Search Tool) searches against NCBI (National Center for Biotechnology Information) Genomes database are also shown (in regular font, starting with "the gi" identification number). The stars highlight the putative BVMO sequences selected for further analysis. Numbers in the nodes correspond to bootstrap values (100 replications, only values higher than 50 are shown).
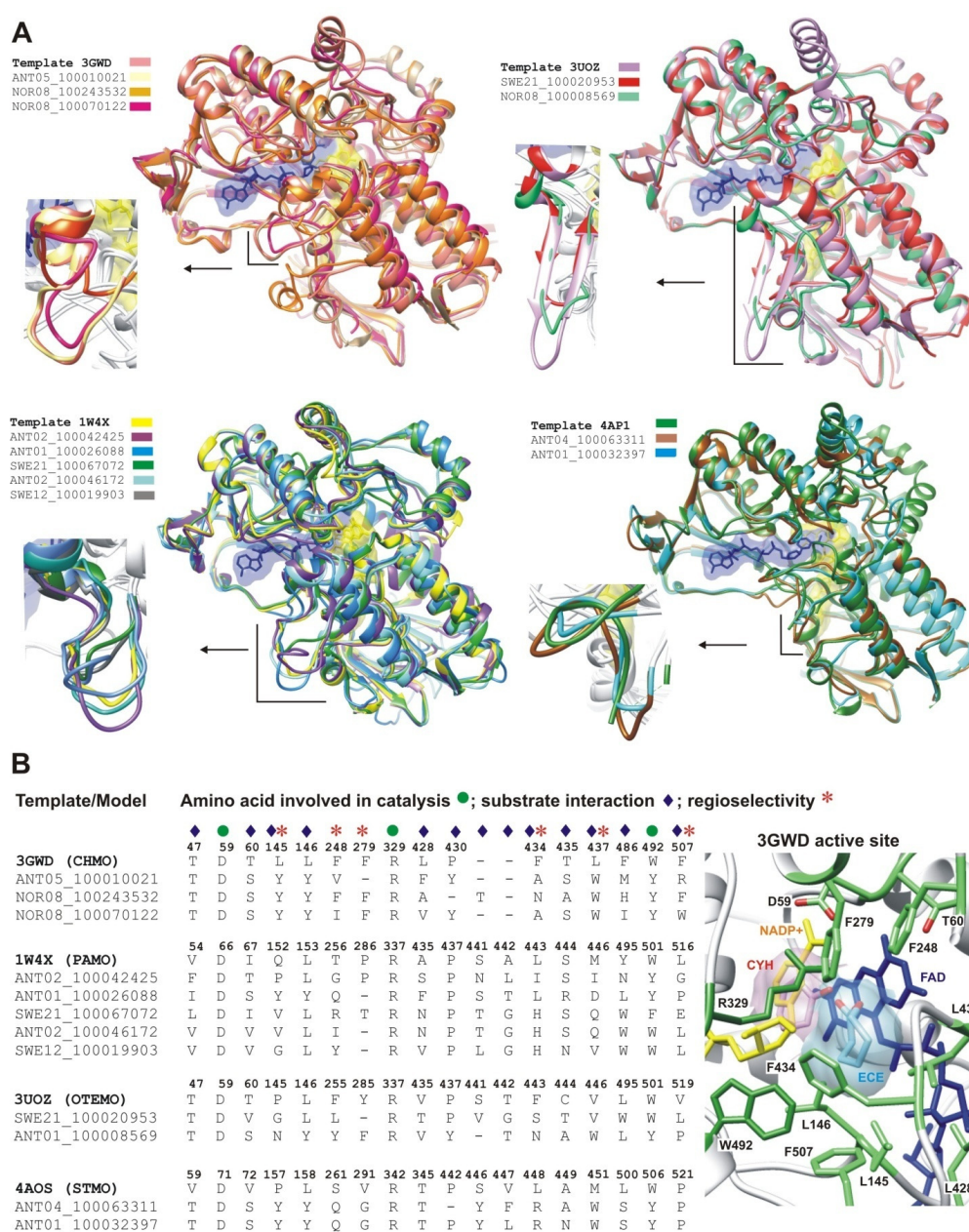
Figure 2. Structural analysis of modeled putative BVMOs. (**A**) structural superimposition between structures that were used as templates for homology modeling and modeled putative BVMOs. FAD (Flavin Adenine Dinucleotide) and NADP$^+$ (Nicotinamide Adenine Dinucleotide Phosphate, oxidized form) as found in template structures are shown in yellow and blue, respectively. Comparisons between Control Loops are highlighted in the insets; (**B**) three-dimensional protein structure analysis indicating amino acid residues involved in catalysis, substrate interaction or regioselectivity. The analysis was derived from structural superimposition in (A), comparing amino acid residues of each template structure with structural models. The numbering corresponds to template structures. The active site of CHMO (Cyclohexanone Monooxygenase) from *Rhodococcus* sp. HI-31 [32] is shown as a reference to indicate the spatial arrangement of key amino acids (green). CYH: substrate cyclohexanone (pink). ECE: product ε-caprolactone (cyan). 3UOZ: OTEMO (2-Oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-Coenzyme A Monooxygenase) from *Pseudomonas putida* ATCC 17453 [33], 1W4X: PAMO (Phenylacetone Monooxygenase) from *Thermobifida fusca* [34], 4AOS: STMO (Steroid Monooxygenase) from *Rhodococcus rhodochrous* [30].

2.2.2. Structural Analysis of the Identified BVMOs: Substrate Range

Among the 12 modeled putative metagenomic BVMOs, four (ANT05_100010021, NOR08_100243532, ANT01_10026088 and SWE21_10067072) showed potential structural features different from the template structures (Figure 3). Lesser steric impediments for the accessibility of the substrate to the catalytic FAD prosthetic group were observed in the modeled metagenomic BVMOs compared with their respective templates structures (Figure 3A,B), even including other crystallized BVMOs such as 2-Oxo-Δ^3-4,5,5-trimethylcyclopentenylacetyl-Coenzyme A monooxygenase (OTEMO) and steroid monooxygenase (STMO) (Figure 3C). The 3D model of the putative BVMO ANT05_100010021 showed an additional site for accessing the catalytic FAD, which was identified as a potential protein channel by the software Channel Finder [35] (Figure 3D). In addition, the binding pockets of these putative enzymes were larger than those of the template BVMOs (Figure 3E).
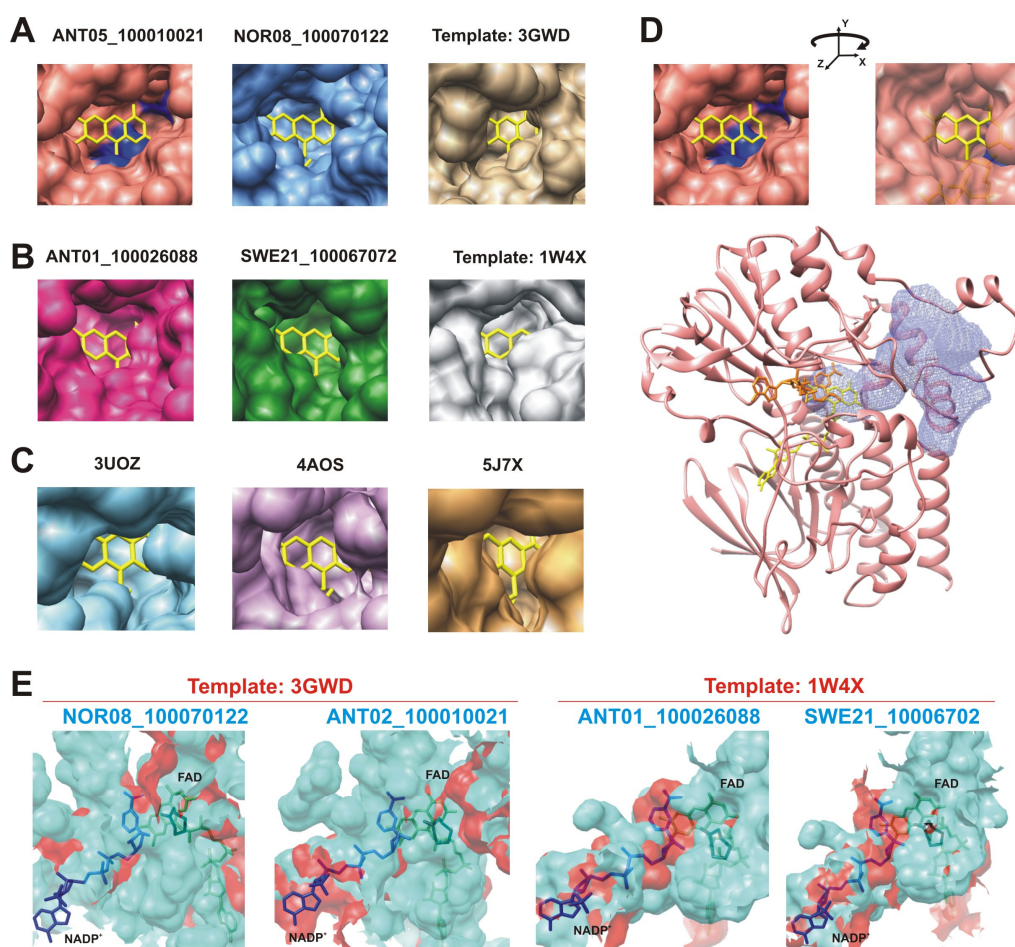
Figure 3. Structural features of four metagenomic putative BVMOs. (**A**) accessibility to the catalytic site of the modeled putative BVMOs ANT05_100010021 and NOR08_100070122 compared with template structure 3GWD; (**B**) comparison of the accessibility to the catalytic site in modeled ANT01_100026088 and SWE21_100067072 with template structure 1W4X; (**C**) accessibility to the catalytic site observed in three additional crystallized BVMOs. 3UOZ: OTEMO from *Pseudomonas putida*; 4AOS: STMO from *Rhodococcus rhodochrous* and 5J7X: BVMO from *Aspergillus flavus*; (**D**) additional site (top, right panel) and protein channel conducting to the catalytic pocket (bottom, light blue surface) identified in the model of ANT05_100010021; (**E**) catalytic pockets of the respective template enzyme (red) compared with the modeled putative BVMOs (cyan). The product ε-caprolactone, such as bound in the crystal structure of CHMO from *Rhodococcus* sp. HI-31 (pdb 4RG3), is shown along with NADP$^+$ and FAD. The active sites were identified with CASTp [36] and identified with Chimera [37].

The potential capability of the metagenomic BVMOs to bind different substrates was assayed through molecular-docking analyses in order to evaluate whether the potential broader substrate accessibility and larger catalytic pockets observed in the models could influence their substrate range (Figure 4). The orientation of the substrate cyclohexanone, such as those bound in the crystal structure of cyclohexanone monooxygenase (CHMO) from *Rhodococcus* sp. HI-31 (pdb 3UCL), was used as the reference to establish the potentially productive orientations of the ligands for catalysis [31].
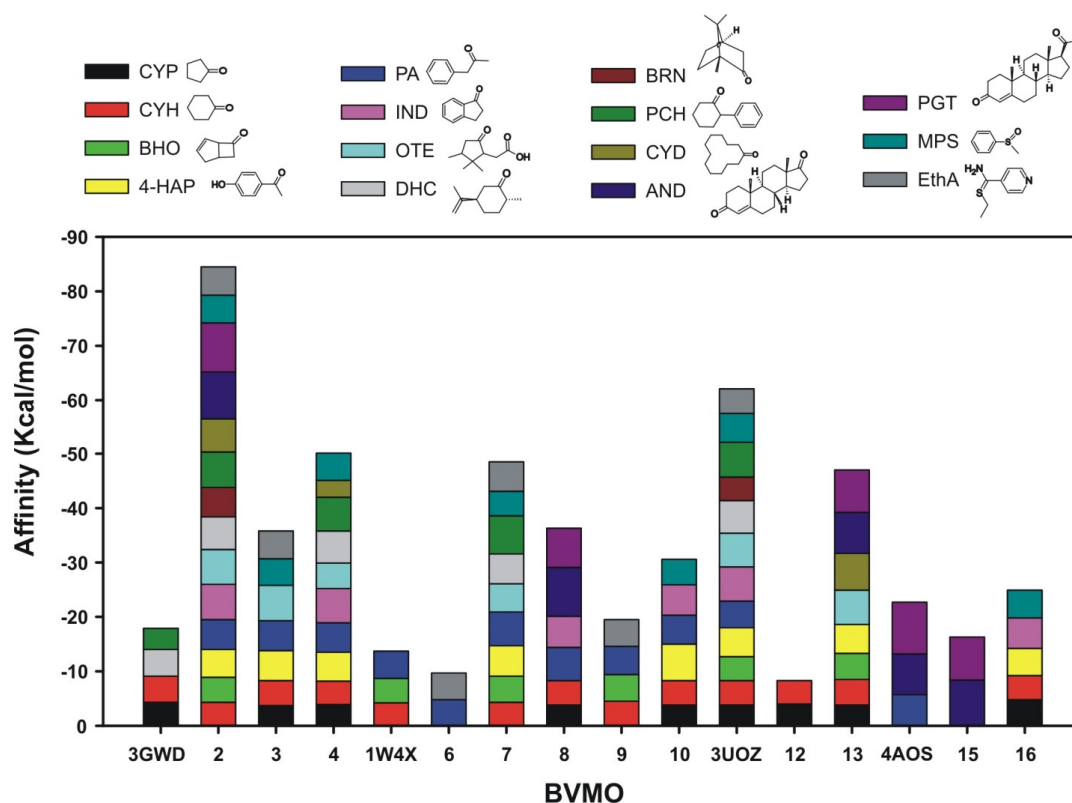
Figure 4. Substrate affinity (kcal/mol) of modeled and crystallized BVMOs estimated by docking analysis. Analyzed substrates are shown at the top. CYP: Cyclopentanone; CYH: Cyclohexanone; BHO: (±)-cisBicyclo[3.2.0]hept-2-en-6-one; 4-HAP: 4-Hydroxyacetophenone; PA: Phenylacetone; IND: Indanone; OTE: 2-Oxo-delta(3)-4,5,5-trimethylcyclopentenylacetic acid; DHC: (1S,4S)-Dihydrocarvone; BRN: Bornanone; PCH: 2-Phenylcyclohexanone; CYD: Cyclododecanone; AND: Androstenedione; PGT: Progesterone; MPS: Methylphenylsulfoxide; Eth: Ethinoamide. 3GWD: CHMO from *Rhodococcus* sp. HI-31 [32]; 2: ANT05_100010021; 3: NOR08_100243532; 4: NOR08_100070122; 1W4X: PAMO from *Thermobifida fusca* [34]; 6: ANT02_100042425; 7: ANT01_100026088; 8: SWE21_100067072; 9: ANT02_100046172; 10: SWE12_100019903; 3UOZ: OTEMO from *Pseudomonas putida* [33]; 12: SWE21_100020953; 13: ANT01_100008569; 4AOS: STMO from *Rhodococcus rhodochrous* [30]; 15: ANT04_100063311; 16: ANT01_100032397. 3D structures of substrate molecules were obtained from PubChem database (compound identification numbers are detailed in Table S3).

Docking analysis suggested that three modeled putative BVMOs (ANT05_100010021, NOR08_100070122 and ANT01_100026088) could recognize 14, 10 and nine different substrates of the 15 analyzed, with ANT05_100010021 displaying the broadest substrate range (Figure 4). The docked orientations of the substrates were productive for catalysis (Figure 5A).
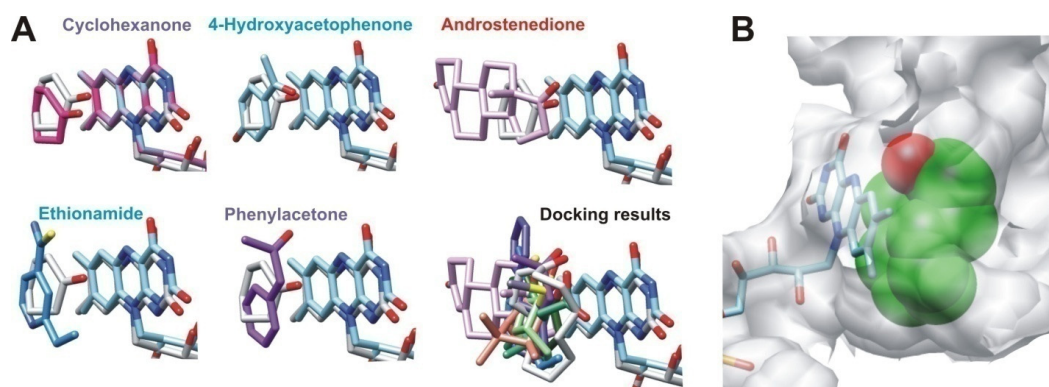
Figure 5. Substrate profile of ANT05_100010021 estimated by molecular-docking analysis. (**A**) obtained spatial arrays, showing results for representative ligands. The results are compared with the geometry adopted by the substrate cyclohexanone in the crystal structure of CHMO from *Rhodococcus* sp. HI-31 (in gray); (**B**) binding of a bulky ligand such as cyclododecanone in the catalytic pocket of the modeled structure.

The additional site that conducts to the catalytic FAD observed in this putative BVMO could be a route for some substrates to enter the catalytic site, broadening the substrate profile. A similar funnel-shaped cavity leading to the active site, observed in phenylacetone monooxygenase (PAMO) from *Thermobifida fusca* [38], supports this hypothesis. In CHMO and OTEMO, this route is blocked by a dipeptide corresponding to residues 278-279 (3GWD CHMO numbering) that is missing in PAMO [31,38]. In the model of ANT05_100010021, there are no residues in a position homologous to this dipeptide (Figure 2B). In addition, the wide catalytic pocket displayed by the ANT05_100010021 model suggests that it could accommodate larger ligands, such as cyclododecanone, broadening its substrate range (Figure 5B).

2.2.3. Structural Flexibility

The substrate entry path in BVMO enzymes is located at the interface between NADP$^+$ and FAD binding domains, where conformational flexibility is high. This flexibility plays an important role in the reshaping of the active site in the presence of different substrates, and has been identified as a structural factor involved in the broad substrate profile displayed by OTEMO BVMOs [33]. Considering the important role of this structural flexibility for catalysis and substrate specificity, the interaction between the FAD and the NADP$^+$ binding domains was analyzed. Structural parameters such as putative ΔG^{diss} and ΔG^{int} were calculated (Table 2).

Table 2. Structural features of modeled BVMOs (Baeyer-Villiger Monooxygenases) and their respective templates.

Enzyme	ΔG^{int} [a] (kcal/mol)	ΔG^{diss} [b] (kcal/mol)	N_{HB} [c]	N_{SB} [d]
3GWD	−30.8	35.2	38	6
NOR08_100243532	−26.5	22.6	21	0
ANT05_100010021	−22.1	17.7	18	5
NOR08_100070122	−28.1	23.5	18	3
1W4X	−51.3	28.3	49	5
ANT01_100026088	−17.7	10.7	13	4
SWE21_100067072	−24.1	18.3	22	7
3UOZ	−30.5	33.2	35	4

[a] ΔG^{int}: Solvation free energy gain upon formation of the interface. Negative values correspond to positive protein affinity; [b] ΔG^{diss}: Free energy of dissociation. Assemblies with $\Delta G^{diss} > 0$ are thermodynamically stable; [c] N_{HB}: Number of potential hydrogen bonds across the interface; [d] N_{SB}: Number of potential salt bridges across the interface; 3GWD: CHMO from *Rhodococcus* sp. HI-31; 1W4X: PAMO from *Thermobifida fusca*; 3UOZ: OTEMO from *Pseudomonas putida*.

The parameter ΔG^{diss} corresponds to the free energy difference between hypothetical dissociated and associated states of FAD and NADP$^+$ binding domains. Positive values of ΔG^{diss} indicate that an external driving force should be applied in order to dissociate the assembly (i.e., the native enzyme); therefore, assemblies with more positive ΔG^{diss} values are thermodynamically more stable. The parameter ΔG^{int} indicates the solvation free energy gain upon formation of the assembly and is restricted to the hydrophobic interactions across the interfaces between domains [39]. Both parameters showed lower interaction at the interface in the metagenomic models than in the templates (Table 2). In addition, fewer numbers of hydrogen and salt bridges between FAD and NADP$^+$ domains were observed in the modeled putative monooxygenases than in their templates. It has been proposed that the formation and disruption of hydrogen-bonding interactions are relevant for the structural flexibility needed to form the catalytic intermediate in BVMO enzymes [31,34].

The results suggest a higher structural flexibility in the vicinity of the active-site pocket of the metagenomic putative BVMOs (Table 2). This feature could render a broader substrate range, although to the detriment of their structural stability, as indicated by the thermodynamic parameters. The wider substrate accessibility and structural flexibility estimated in the metagenomic putative BVMOs could be a result of adaptation to cold temperatures, as proposed for psychrophilic enzymes [40].

### 2.2.4. Structural Analysis of Identified BVMOs: Substrate Regioselectivity

The potential regioselectivity of the three putative BVMOs modeled using as template the crystal structure of CHMO from *Rhodococcus* sp. HI-31 bound to the product ε-caprolactone in the tight conformation (pdb 4RG3) (ANT05_100010021, NOR08_100243532 and NOR08_100070122) was assayed by molecular docking analysis. This structure has been proposed to be a suitable scaffold for studying enantio- or regioselectivity [28,41]. The oxidation of (+)-*trans*-dihydrocarvone was used as a model reaction, according to previous studies reported in the literature [28]. This analysis was limited to these putative enzymes because, up to now, the CHMO from *Rhodococcus* sp. HI-31 is the only BVMO for which the structural information necessary for estimating supported product bindings is available. Two regioisomers are possible products of the oxidation of alpha-substituted ketones substrates by a BVMO, the normal lactone formed by insertion of the oxygen atom next to the more substituted carbon and the abnormal lactone formed when the oxygen atom is inserted next to the less substituted carbon. The regioselective oxidation of (+)-*trans*-dihydrocarvone by a set of BVMOs showed that most CHMOs, including the CHMO from *Rhodococcus* sp. HI-31, produce the abnormal lactone only [42]. However, a complete switch of regioselectivity of the CHMO from *Arthrobacter* sp. BP2 toward (+)-*trans*-dihydrocarvone was achieved by site-directed mutagenesis on relevant positions, leading to the normal lactone [28]. The analysis presented in this work suggests that putative BVMOs NOR08_100243532 and NOR_100070122 would be regioselective for abnormal lactones, but NOR08_100243532 would be specific for small molecules (Table 3).

These results suggest that ANT05_100010021 would be the only putative BVMO able to produce normal lactone, even from substrates smaller or larger than (+)-*trans*-dihydrocarvone (Table 3). However, this putative enzyme would also be able to produce the abnormal lactones (Table 3), to the detriment of regioselectivity. The presence of gaps or small side-chain amino acids in the modeled ANT05_100010021 instead of bulkier residues typically observed in wild-type CHMOs could prevent steric clashes with lactone products, with the concomitant capability of binding normal lactone (Figure 6). A similar mechanism was proposed for the triple variant of CHMO from *Arthrobacter* sp. BP2 in which the mutation of F248, F279 and F434 to alanine completely switched the regioselectivity of the enzyme and produced the normal lactone [28]. The observed differences in these residues in the structural model of ANT05_100010021 (valine instead of F248, gap instead of F279 and alanine instead of F434) suggest that normal lactone could be produced (Figure 6). However, the resultant wide space in the catalytic cavity could preclude regioselectivity.

Table 3. Affinities of modeled putative BVMOs and CHMO (Cyclohexanone Monooxygenase) from *Rhodococcus* sp. HI-31 by products as estimated by docking analysis (kcal/mol).

Structure of Lactone	BVMO			
Products	ANT05_100010021	NOR08_100243532	NOR08_100070122	4RG3
A- B- (Normal)	−4.5	-4.7	-5.0	-6.0
C- (Abnormal)	−5.3	NB	NB	NB
D- (Normal)	−5.1	-5.0	-5.5	-6.2
E- (Abnormal)	−6.1	NB	NB	NB
F- (Normal)	−5.8	NB	-5.9	-4.5
G- (Abnormal)	−6.9	NB	NB	NB
	−4.9	NB	-6.2	NB

A: ε-Caprolactone; B: 3-Methyloxepan-2-one; C: 7-Methyl-2-oxepanone; D: (4R,7R)-4 isopropenyl-7-methyloxepan-2-one; E: (3R,6R)-6-isopropenyl-3-methyloxepan-2-one; F: 10-methyloxecan-2-one; G: 3-methyloxecan-2-one. NB: No binding. 4RG3: CHMO from *Rhodococcus* sp. HI-31 bound to ε-caprolactone [41]. 3D structures of product molecules were obtained from PubChem database (compound identification numbers are detailed in Table S3).
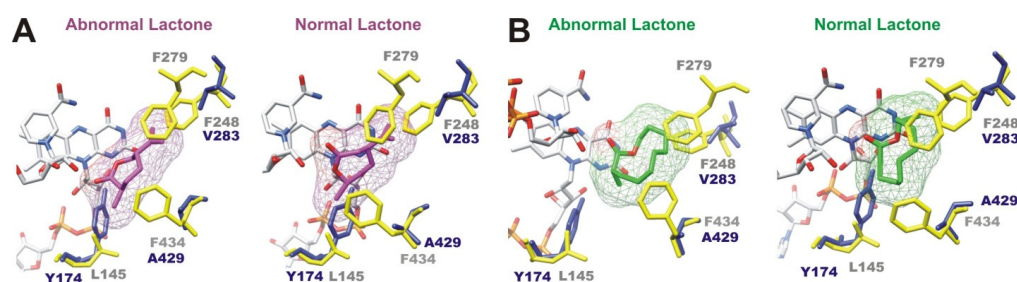
Figure 6. In silico regioselectivity assays of modeled ANT05_100010021 by docking analysis. (**A**) docking results of ANT05_100010021 (blue) with lactone products derived from the oxidation of (+)-*trans*-dihydrocarvone (magenta, mesh representation). Abnormal lactone: (3R,6R)-6-isopropenyl-3-methyloxepan-2-one. Normal lactone: (4R,7R)-4 isopropenyl-7-methyloxepan-2-one; (**B**) docking results of ANT05_100010021 (blue) with lactone products derived from the oxidation of 2-methyl-cyclodecanone (green, mesh representation). Abnormal lactone: 3-methyloxecan-2-one. Normal lactone: 10-methyloxecan-2-one. Only key residues involved in regioselectivity are shown. Structural superimpositions between modeled ANT05_100010021 (blue) and template BVMO (yellow, CHMO from *Rhodococcus* sp. HI-31 in the tight conformation, pdb 4RG3) show differences in amino acids involved in regioselectivity.

2.3. Cytochrome P450 CYP153

The cytochrome P450 family comprises a wide variety of enzymes including CYP153, which are able to hydroxylate medium-long alkanes (C_5 to C_{16}) at the terminal carbon [43]. Out of the 156 PCS for putative cytochrome P450 identified in the metagenomic dataset, nine clustered with sequences from characterized CYP153 enzymes (Figure S2). Two of these sequences grouped with the phylogenetic cluster II defined by Nie and collaborators [20] and the other seven with the cluster IV. Only the metagenomic sequences ARG05_10097442 (cluster II), ANT06_10083082 and ANT06_10083083 (cluster IV) were full-length, and were modeled using the crystal structure of the cytochrome P450pyr hydroxylase from *Sphingomonas* sp. HXN-200 (pdb 3RWL) [44]. The application of this enzyme for the synthesis of (S)-N-benzyl 3-hydroxy-pyrrolidine has been reported [44], which is a useful intermediate for the preparation of several pharmaceutical products, antibiotic drugs and agricultural chemicals [45]. To improve its performance as a biocatalyst, this enzyme has been engineered by iterative saturation mutagenesis and the critical amino acids for sterero-specificity were identified [44,46]. These residues surround the catalytic haem and are critical not only for sterero-specificity but also for substrate specificity. The 3D model of ARG05_10097442 is shown in Figure 7A, and similar results were obtained with other putative CYP153 enzyme sequences (data not shown).
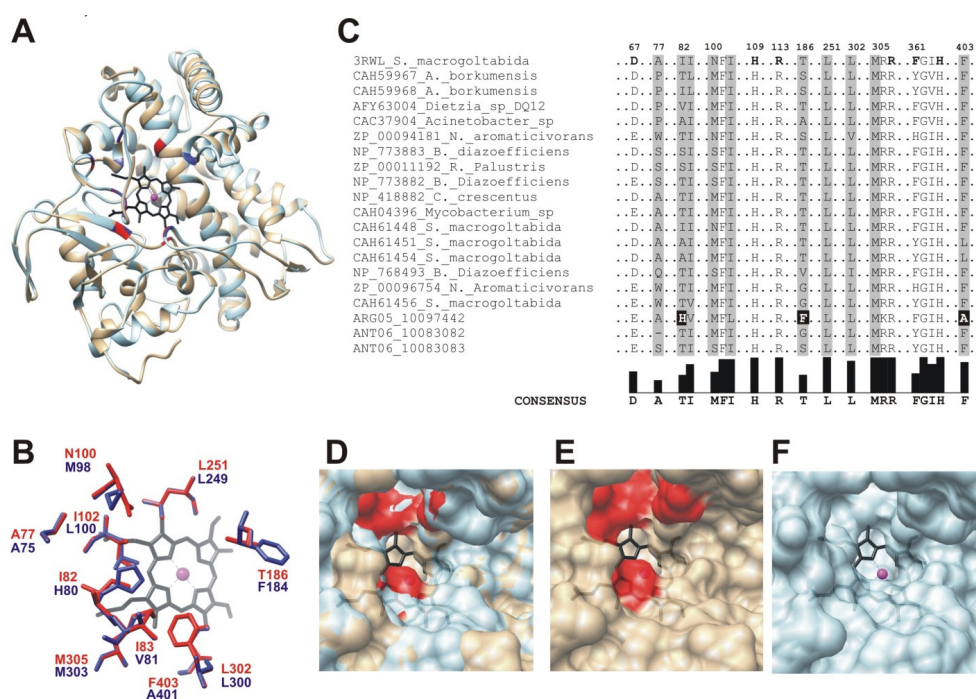
Figure 7. Structural features of a metagenomic putative cytochrome P450 CYP153. (**A**) structural superimposition between the modeled ARG05_10097442 (light blue) and template (grey) CYP153. The haem prosthetic group is shown in black in the center of the structure, with the iron atom as magenta sphere; (**B**) alignment of the characterized CYP153 and the metagenomic sequences. Amino acids involved in haem interaction are in bold. The amino acids relevant for stereo-specificity are in grey boxes, with the number relatives to the crystal structure of CYP153 from *Sphingopyxis macrogoltabida* (pdb 3RWL). As shown in white, the sequence ARG05_10097442 displays unique amino acid modifications; (**C**) spatial arrangement of the amino acids relevant for stereo-specificity (highlighted in grey boxes in Figure 5B) of the template (blue, 3RWL), compared with the model (red); (**D**) surface representation of the overlapped active sites of the templates and model; (**E**) active site of the template (3RWL), showing in red the amino acids relevant for stereo-specificity; (**F**) active site of the modeled ARG05_10097442.

Different residues were observed in the model with respect to template structure at amino acid residues relevant for stereo-specificity and substrate specificity (Figure 7B). In order to determine if these amino acid differences are present in other CYP153s, the sequences identified in this study were compared with characterized CYP153s [20] (Figure 7C). The sequence ARG05_10097442 showed differences in the positions of the amino acids relevant for stereo-specificity (black boxes, Figure 7C), which were not found by saturation mutagenesis of P450pyr hydroxylase from *Sphingomonas* sp. HXN-200 [44,46]. The residues observed in the sequence ARG05_10097442 could reduce steric impairments around the catalytic haem (Figure 7D–F), allowing changes in catalytic properties related to substrate/product features and probably a novel substrate/product stereo-specificity.

3. Materials and Methods

3.1. Characteristic of Cold Sediments' Metagenomic Dataset

The metagenomic dataset analyzed in this study was obtained by shotgun sequencing of DNA isolated from 23 sediment samples retrieved from four high-latitude coastal environments: (i) Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway [NOR]; (ii) Port Värtahamnen, Stockholm, Baltic Sea, Sweden [SWE]; (iii) Ushuaia Bay, Tierra del Fuego Island, Argentina [ARG]; and (iv) Potter Cove, 25 de Mayo (King George) Island, Antarctica [ANT]. Details on sampling, DNA extraction and shotgun sequencing were previously reported [21]. Briefly, shotgun sequencing was performed using Illumina HiSeq 1500 (2 × 150-bp paired end reads, one lane per sample), at the facilities of the Joint Genome Institute, Department of Energy, 2800 Mitchell Drive, Walnut Creek, CA, USA. The 23 metagenomes, including unassembled reads and scaffolds, were annotated using the IMG (Integrated Microbial Genomes) pipeline [47]. This dataset contains a total of 13,931,912 CDS (coding sequences) in the assembled fraction [21]. The sequences are available at the IMG/M (Integrated Microbial Genomes and Microbiomes) server [47] under accession numbers 3300000118-3300000136, 3300000241-3300000243, and 3300000792.

3.2. Screening of the Metagenomic Dataset

For each enzyme family (BVMO and CYP153), sequences were obtained from the metagenomes using two complementary functional evidences [47]. Firstly, sequences containing Pfam domains [48] PF00067 ("Cytochrome P450") or PF00743 ("Flavin-binding monooxygenase-like") were selected and downloaded from the IMG server. Secondly, BLASTP searches with a cut-off E-value of 10^{-5} were performed in IMG, using selected sequences from crystallized and/or biochemically characterized enzymes as query (Table S1). Duplicated sequences were eliminated and full-length or nearly full-length sequences were retrieved from the datasets by preselecting sequences longer than 250 or 300 amino acids, depending on the enzyme (Table S1). An additional filtering was applied to metagenomic P450 sequences, as the P450 superfamily included more than 1000 families of bacteria alone [49]. Therefore, a representative set of 1,122 bacterial P450 sequences was downloaded from the bacterial cytochrome P450 database [49], and only sequences whose first match in a standalone blast analysis corresponded to CYP153 family members in the database were kept.

3.3. Phylogenetic Analysis

Reference sequences for each enzyme family were collected from the literature. Sequences coding for cytochrome P450 were obtained from a recent work published by Nie et al. [20]. For BVMOs, an in-house-built database was constructed from sequences previously selected by Huijbers et al. and Mascotti et al. [14,24]. In addition, BLASTP first matches of the obtained metagenomic sequences against NCBI (National Center for Biotechnology Information, Bethesda, MD, USA) Representative Genomes database were also added to the analysis, as many of the sequences were found to be divergent from those chosen as reference. Reference sequences were aligned in ClustalX [50] with default parameters, followed by multiple alignment of the identified metagenomic sequences and their

first matches against the reference alignment, using ClustalX. Sequence alignments were manually trimmed in order to obtain an equal number of positions, and maximum-likelihood phylogenetic trees were constructed in RAxML version 8.2.3 (Heidelberg Institute for Theoretical Studies, D-69118 Heidelberg, Germany) [51], with GAMMA model of rate heterogeneity, LG substitution matrix and empirical base frequencies (option PROTGAMMALGF). Bootstrapping was performed with 100 replications.

3.4. Three-Dimensional Protein Structure Modeling and Model Quality Evaluation

For each metagenomic sequence, the HHpred web server was used to search for suitable templates for building high quality models [52]. Selected templates presented the highest sequence identity, coverage and resolution (Table S2). For each metagenomic sequence, five models were calculated using MODELLER version 9.16 (Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Byers Hall, University of California San Francisco, San Francisco, CA, USA) [53]. Models were ranked using DOPE (Discrete Optimized Protein Energy) Z scores. The top models with the lowest energy scores for each metagenomic sequence were further validated by Verify3D [54] and PROCHECK [55]. A final step of refinement was performed with the server 3Drefine [56]. Clashes remotion was carried out using Chimera [37] and Swiss PDB (Protein Data Bank files) viewer [57]. Energy minimization was performed with YASARA (Yet Another Scientific Artificial Reality Application) [58].

3.5. Docking Analysis

Three-dimensional structure files of ligand molecules were downloaded from the PubChem database [59], and compound identifier numbers (CID) are detailed in Table S3. Docking analyses were carried out using the software AutoDock Vina (Molecular Graphics Lab at The Scripps Research Institute, La Jolla, CA, USA) with an exhaustiveness of 8.0 [60]. Modeled structures were superimposed to template structures co-crystallized with substrates (or products) in functional orientation. The different ligands to be in silico assayed were manually superimposed to the substrate (or product) bound to the active site. The enzyme and ligand molecules were saved separately as different PDB files and further loaded in AutoDock Vina. Grid parameters were set using coordinates comprising the active site. Only affinity values corresponding to geometry bindings compatible for catalysis were considered.

3.6. Calculation of Structural Parameters

Parameters associated with protein structure stability were calculated using PDBePISA server (Protein Interfaces, Surfaces and Assemblies) [39]. For calculations, FAD and NADP$^+$ binding subdomains were considered as different interacting protein chains. The information to define subdomains was obtained from the crystal structure of the different Baeyer–Villiger Monooxygenase used as templates [33]. For the modeled enzymes, the same space symmetry groups of the crystal structures used as templates were considered. The catalytic pockets were calculated with the CASTp server using a probe radius of 1.4 Å [36]. The obtained poc files were analyzed with the software chimera [37]. The identification of protein channels was carried out with the software Channel Finder (The Scripps Research Institute and Program in Computational Biology and Bioinformatics New Haven, CT, USA) of the Web server 3V, using 10 and 3 as outer and inner probe radius, respectively [35].

4. Conclusions

The methodological framework reported in this work allowed the selection of four sequences out of ~14 million PCSs as candidates for synthesis and heterologous expression, greatly aiding in the efficient and knowledge-based exploitation of a highly fragmented metagenomic dataset. These sequences include the putative BVMOs ANT05_100010021, NOR08_100070122 and ANT01_100026088, which were predicted to have a wide substrate range and novel substrate specificities. In addition,

NOR08_100070122 and NOR08_100243532 could be regioselectives for abnormal lactones. On the other hand, the putative CYP153 ARG05_10097442 could present novel substrate specificity and regioselectivity. These putative enzymes could be used in the synthesis of a wide variety of lactones or 1-alkanols at room or low temperatures, where biooxidations may produce better results than the typical chemical oxidation with regard to performance and cost savings in energy consumption. These putative enzymes also could be applied to improve procedures aimed to overcome substrate and product inhibition, performing the reaction in a biphasic system by using organic solvents whose water miscibility can be varied by temperature [19]. Considering that the enzymes could catalyze the reaction at low temperatures, upon this condition, the solubility of the substrate in the aqueous phase would be diminished, improving the feature of the organic phase to trap the substrate during the reaction and favoring the extraction of the product from the aqueous phase.

The presented evidence suggests that the active sites of the selected putative monooxygenases would be larger and more accessible than those from previously biochemically characterized enzymes. It has been observed that low temperature adapted enzymes present larger catalytic cavities, more accessible to ligands, than mesophilic enzymes [40]. A better accessibility is suggested not only to be responsible for the accommodation of the substrate at low energy cost but also to facilitate the release and exit of the reaction products [40]. Structural flexibility also plays a relevant role for cold adaptation because it involves a tuned destabilization of the active site or the whole protein, allowing the catalytic center to be more mobile or flexible at temperatures that tend to freeze molecular motions [40]. In BVMOs, the structural plasticity represents an advantage for catalysis considering the proposed need of these enzymes to bind cofactors and substrates with diverse structures in order to catalyze the elaborated chemical mechanism [31]. These features were observed in the modeled enzymes, suggesting structural mechanisms for cold adaptation. These structural features could favour oxygenation reactions at low or medium temperature with a wide range of substrates. Moreover, novel stereo-specificities could be potentially obtained as well, considering that BVMOs are continuously evolving to acquire new activities, depending on the emerging availabilities of new compounds in the natural environment [30].

Finally, the methodology applied in this work can be used for mining genetic information with biotechnological potential from metagenomic datasets currently stored in public databases, as well as to complement approaches based on functional metagenomics or where only specific HMMs are constructed and applied in screening. This complement may strengthen the prediction aimed to rationale biocatalysts selection before biochemical characterization.

Author Contributions: H.M.D., J.C., J.K.J., S.S. and W.P.M.C. designed sampling experiments and sampled the sediments; M.A.M., D.V.R., M.L. and H.M.D. analyzed the data; M.L. performed the phylogenetic analysis; M.A.M. performed screening of the metagenomic dataset, protein modeling, docking analysis and calculation of structural parameters; and M.A.M. and H.M.D. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Trincone, A. Some enzymes in marine environment: Prospective applications found in patent literature. *Recent Patents Biotechnol.* **2012**, *6*, 134–148. [CrossRef]

2. Lozada, M.; Dionisi, H.M. Microbial bioprospecting in marine environments. In *Springer Handbook of Marine Biotechnology*; Kim, S.K., Ed.; Springer: Berlin, Germany, 2015; pp. 307–326.

3. Delmont, T.O.; Simonet, P.; Vogel, T.M. Describing microbial communities and performing global comparisons in the omic era. *ISME J.-Int. Soc. Microb. Ecol.* **2012**, *6*, 1625–1628. [CrossRef] [PubMed]

4. Holmes, A.J.; Coleman, N.V. Evolutionary ecology and multidisciplinary approaches to prospecting for monooxygenases as biocatalysts. *Antonie Leeuwenhoek* **2008**, *94*, 75–84. [CrossRef] [PubMed]

5. Ferrer, M.; Martinez-Martinez, M.; Bargiela, R.; Streit, W.R.; Golyshina, O.V.; Golyshin, P.N. Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microb. Biotechnol.* **2015**, *9*, 22–34. [CrossRef] [PubMed]

6. Currin, A.; Swainston, N.; Day, P.J.; Kell, D.B. Synthetic biology for the directed evolution of protein biocatalysts: Navigating sequence space intelligently. *Chem. Soc. Rev.* **2015**, *44*, 1172–1239. [CrossRef] [PubMed]

7. Holtmann, D.; Fraaije, M.W.; Arends, I.W.; Opperman, D.J.; Hollmann, F. The taming of oxygen: Biocatalytic oxyfunctionalisations. *Chem. Commun.* **2014**, *50*, 13180–13200. [CrossRef] [PubMed]

8. Yang, G.; Ding, Y. Recent advances in biocatalyst discovery, development and applications. *Bioorg. Med. Chem.* **2014**, *22*, 5604–5612. [CrossRef] [PubMed]

9. Pazmino, D.T.; Winkler, M.; Glieder, A.; Fraaije, M. Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *J. Biotechnol.* **2010**, *146*, 9–24. [CrossRef] [PubMed]

10. Li, Z.; van Beilen, J.B.; Duetz, W.A.; Schmid, A.; de Raadt, A.; Griengl, H.; Witholt, B. Oxidative biotransformations using oxygenases. *Curr. Opin. Chem. Biol.* **2002**, *6*, 136–144. [CrossRef]

11. Bučko, M.; Gemeiner, P.; Schenkmayerová, A.; Krajčovič, T.; Rudroff, F.; Mihovilovič, M.D. Baeyer-villiger oxidations: Biotechnological approach. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 6585–6599. [CrossRef] [PubMed]

12. Bernhardt, R.; Urlacher, V.B. Cytochromes p450 as promising catalysts for biotechnological application: Chances and limitations. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 6185–6203. [CrossRef] [PubMed]

13. Ji, Y.; Mao, G.; Wang, Y.; Bartlam, M. Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Front. Microbiol.* **2013**, *4*, 58. [CrossRef] [PubMed]

14. Huijbers, M.M.; Montersino, S.; Westphal, A.H.; Tischler, D.; van Berkel, W.J. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **2014**, *544*, 2–17. [CrossRef] [PubMed]

15. Ceccoli, R.D.; Bianchi, D.A.; Rial, D.V. Flavoprotein monooxygenases for oxidative biocatalysis: Recombinant expression in microbial hosts and applications. *Recomb. Protein Expr. Microb. Syst.* **2014**, *5*, 25. [CrossRef] [PubMed]

16. Alphand, V.; Wohlgemuth, R. Applications of baeyer-villiger monooxygenases in organic synthesis. *Curr. Org. Chem.* **2010**, *14*, 1928–1965. [CrossRef]

17. Fraaije, M.W.; Janssen, D.B. Biocatalytic scope of baeyer-villiger monooxygenases. In *Modern Biooxidation: Enzymes, reaction and applications*; Schmid, R.D., Urlacher, V.B., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA.: Darmstadt, Germany, 2007; pp. 77–97.

18. Bong, Y.K.; Collier, S.J.; Mijts, B.; Vogel, M.; Zhang, X.; Zhu, J.; Nazor, J.S.D.; Song, S. Synthesis of Prazole Compounds. WO2011071982 A2, 2011.

19. Leisch, H.; Morley, K.; Lau, P.C. Baeyer-villiger monooxygenases: More than just green chemistry. *Chem. Rev.* **2011**, *111*, 4165–4222. [CrossRef] [PubMed]

20. Nie, Y.; Chi, C.-Q.; Fang, H.; Liang, J.-L.; Lu, S.-L.; Lai, G.-L.; Tang, Y.-Q.; Wu, X.-L. Diverse alkane hydroxylase genes in microorganisms and environments. *Sci. Rep.* **2014**, *4*, 4968. [CrossRef] [PubMed]

21. Matos, M.N.; Lozada, M.; Anselmino, L.E.; Musumeci, M.A.; Henrissat, B.; Jansson, J.K.; Mac Cormack, W.P.; Carroll, J.; Sjöling, S.; Lundgren, L. Metagenomics unveils the attributes of the alginolytic guilds of sediments from four distant cold coastal environments. *Environ. Microbiol.* **2016**, *18*, 4471–4484. [CrossRef] [PubMed]

22. Raddadi, N.; Cherif, A.; Daffonchio, D.; Neifar, M.; Fava, F. Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7907–7913. [CrossRef] [PubMed]

23. Howe, A.C.; Jansson, J.K.; Malfatti, S.A.; Tringe, S.G.; Tiedje, J.M.; Brown, C.T. Tackling soil diversity with the assembly of large, complex metagenomes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4904–4909. [CrossRef] [PubMed]

24. Mascotti, M.L.; Lapadula, W.J.; Ayub, M.J. The origin and evolution of baeyer-villiger monooxygenases (bvmos): An ancestral family of flavin monooxygenases. *PLoS ONE* **2015**, *10*, e0132689. [CrossRef] [PubMed]

25. Fraaije, M.W.; Kamerbeek, N.M.; van Berkel, W.J.; Janssen, D.B. Identification of a baeyer-villiger monooxygenase sequence motif. *FEBS Lett.* **2002**, *518*, 43–47. [CrossRef]

26. Riebel, A.; Dudek, H.; De Gonzalo, G.; Stepniak, P.; Rychlewski, L.; Fraaije, M. Expanding the set of rhodococcal baeyer-villiger monooxygenases by high-throughput cloning, expression and substrate screening. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 1479–1489. [CrossRef] [PubMed]

27. Yachnin, B.J.; Lau, P.C.; Berghuis, A.M. The role of conformational flexibility in baeyer-villiger monooxygenase catalysis and structure. *Biochim. Biophys. Acta* **2016**, *1864*, 1641–1648. [CrossRef] [PubMed]

28. Balke, K.; Schmidt, S.; Genz, M.; Bornscheuer, U.T. Switching the regioselectivity of a cyclohexanone monooxygenase toward (+)-trans-dihydrocarvone by rational protein design. *ACS Chem. Biol.* **2016**, *11*, 38–43. [CrossRef] [PubMed]

29. Dudek, H.M.; de Gonzalo, G.; Pazmiño, D.E.T.; Stępniak, P.; Wyrwicz, L.S.; Rychlewski, L.; Fraaije, M.W. Mapping the substrate binding site of phenylacetone monooxygenase from thermobifida fusca by mutational analysis. *Appl. Environ. Microbiol.* **2011**, *77*, 5730–5738. [CrossRef] [PubMed]

30. Franceschini, S.; van Beek, H.L.; Pennetta, A.; Martinoli, C.; Fraaije, M.W.; Mattevi, A. Exploring the structural basis of substrate preferences in baeyer-villiger monooxygenases insight from steroid monooxygenase. *J. Biol. Chem.* **2012**, *287*, 22626–22634. [CrossRef] [PubMed]

31. Yachnin, B.J.; Sprules, T.; McEvoy, M.B.; Lau, P.C.; Berghuis, A.M. The substrate-bound crystal structure of a baeyer-villiger monooxygenase exhibits a criegee-like conformation. *J. Am. Chem. Soc.* **2012**, *134*, 7788–7795. [CrossRef] [PubMed]

32. Mirza, I.A.; Yachnin, B.J.; Wang, S.; Grosse, S.; Bergeron, H.; Imura, A.; Iwaki, H.; Hasegawa, Y.; Lau, P.C.; Berghuis, A.M. Crystal structures of cyclohexanone monooxygenase reveal complex domain movements and a sliding cofactor. *J. Am. Chem. Soc.* **2009**, *131*, 8848–8854. [CrossRef] [PubMed]

33. Leisch, H.; Shi, R.; Grosse, S.; Morley, K.; Bergeron, H.; Cygler, M.; Iwaki, H.; Hasegawa, Y.; Lau, P.C. Cloning, baeyer-villiger biooxidations, and structures of the camphor pathway 2-oxo-δ3-4, 5, 5-trimethylcyclopentenylacetyl-coenzyme a monooxygenase of pseudomonas putida atcc 17453. *Appl. Environ. Microbiol.* **2012**, *78*, 2200–2212. [CrossRef] [PubMed]

34. Malito, E.; Alfieri, A.; Fraaije, M.W.; Mattevi, A. Crystal structure of a baeyer-villiger monooxygenase. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13157–13162. [CrossRef] [PubMed]

35. Voss, N.R.; Gerstein, M. 3v: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555–W562. [CrossRef] [PubMed]

36. Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. Castp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **2006**, *34*, W116–W118. [CrossRef] [PubMed]

37. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]

38. Orru, R.; Dudek, H.M.; Martinoli, C.; Pazmiño, D.E.T.; Royant, A.; Weik, M.; Fraaije, M.W.; Mattevi, A. Snapshots of enzymatic baeyer-villiger catalysis oxygen activation and intermediate stabilization. *J. Biol. Chem.* **2011**, *286*, 29284–29291. [CrossRef] [PubMed]

39. Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774–797. [CrossRef] [PubMed]

40. Feller, G.; Gerday, C. Psychrophilic enzymes: Hot topics in cold adaptation. *Nat. Rev. Microbiol.* **2003**, *1*, 200–208. [CrossRef] [PubMed]

41. Yachnin, B.J.; McEvoy, M.B.; MacCuish, R.J.; Morley, K.L.; Lau, P.C.; Berghuis, A.M. Lactone-bound structures of cyclohexanone monooxygenase provide insight into the stereochemistry of catalysis. *ACS Chem. Biol.* **2014**, *9*, 2843–2851. [CrossRef] [PubMed]

42. Černuchová, P.; Mihovilovic, M.D. Microbial baeyer-villiger oxidation of terpenones by recombinant whole-cell biocatalysts—Formation of enantiocomplementary regioisomeric lactones. *Org. Biomol. Chem.* **2007**, *5*, 1715–1719. [CrossRef] [PubMed]

43. Van Beilen, J.; Funhoff, E.; Van Loon, A.; Just, A.; Kaysser, L.; Bouza, M.; Holtackers, R.; Rothlisberger, M.; Li, Z.; Witholt, B. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. *Appl. Environ. Microbiol.* **2006**, *72*, 59–65. [CrossRef] [PubMed]

44. Pham, S.Q.; Pompidor, G.; Liu, J.; Li, X.-D.; Li, Z. Evolving P450pyr hydroxylase for highly enantioselective hydroxylation at non-activated carbon atom. *Chem. Commun.* **2012**, *48*, 4618–4620. [CrossRef] [PubMed]

45. Taneja, S.C.; Aga, M.A.; Kumar, B.; Sethi, V.K.; Andotra, S.S.; Qazi, G.N. Process for the Preparation of Optically Active N-Benzyl-3 Hydroxypyrrolidines. US8445700 B2, 2012.

46. Yang, Y.; Li, Z. Evolving P450pyr monooxygenase for regio-and stereoselective hydroxylations. *Chim. Int. J. Chem.* **2015**, *69*, 136–141. [CrossRef] [PubMed]

47. Markowitz, V.M.; Chen, I.-M.A.; Chu, K.; Szeto, E.; Palaniappan, K.; Pillay, M.; Ratner, A.; Huang, J.; Pagani, I.; Tringe, S. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **2014**, *42*, D568–D573. [CrossRef] [PubMed]

48. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, D222–D230. [CrossRef] [PubMed]

49. Kelly, S.L.; Kelly, D.E. Microbial cytochromes p450: Biodiversity and biotechnology. Where do cytochromes p450 come from, what do they do and what can they do for us? *Phil. Trans. R. Soc. B* **2013**, *368*. [CrossRef] [PubMed]

50. Larkin, M.A.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef] [PubMed]

51. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef] [PubMed]

52. Söding, J.; Biegert, A.; Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244–W248. [CrossRef] [PubMed]

53. Webb, B.; Sali, A. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* **2006**, *47*, 5.

54. Eisenberg, D.; Lüthy, R.; Bowie, J.U. Verify3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **1997**, *277*, 396. [PubMed]

55. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291. [CrossRef]

56. Bhattacharya, D.; Cheng, J. 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. *Proteins* **2013**, *81*, 119–131. [CrossRef] [PubMed]

57. Guex, N.; Peitsch, M.C. Swiss-model and the swiss-pdb viewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723. [CrossRef] [PubMed]

58. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **2009**, *77*, 114–122. [CrossRef] [PubMed]

59. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. Pubchem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [CrossRef] [PubMed]

60. Trott, O.; Olson, A.J. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef] [PubMed]