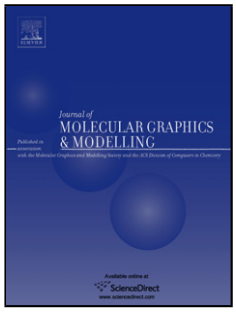


Title: Molecular Modelling and Quantum Biochemistry Computations of a Naturally Occurring Bioremediation Enzyme: Alkane Hydroxylase from *Pseudomonas putida* P1

Authors: B.G. de Sousa, J.I.N. Oliveira, E.L. Albuquerque, U.L. Fulco, V.E. Amaro, C.A.G. Blaha

PII: S1093-3263(17)30458-8
DOI: http://dx.doi.org/10.1016/j.jmgm.2017.08.021
Reference: JMG 7013

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 19-6-2017
Revised date: 24-8-2017
Accepted date: 26-8-2017

Please cite this article as: B.G.de Sousa, J.I.N.Oliveira, E.L.Albuquerque, U.L.Fulco, V.E.Amaro, C.A.G.Blaha, Molecular Modelling and Quantum Biochemistry Computations of a Naturally Occurring Bioremediation Enzyme: Alkane Hydroxylase from Pseudomonas putida P1, Journal of Molecular Graphics and Modellinghttp://dx.doi.org/10.1016/j.jmgm.2017.08.021

Molecular Modelling and Quantum Biochemistry Computations of a Naturally Occurring Bioremediation Enzyme: Alkane Hydroxylase from *Pseudomonas putida* P1

B. G. de Sousa[1*], J. I. N. Oliveira[2], E. L. Albuquerque[2], U. L. Fulco[2], V. E. Amaro[1,3] and C. A. G. Blaha[4]

[1] *Programa de Pós-graduação em Ciência e Engenharia de Petróleo, Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte, 59072-970 Natal-RN, Brazil*

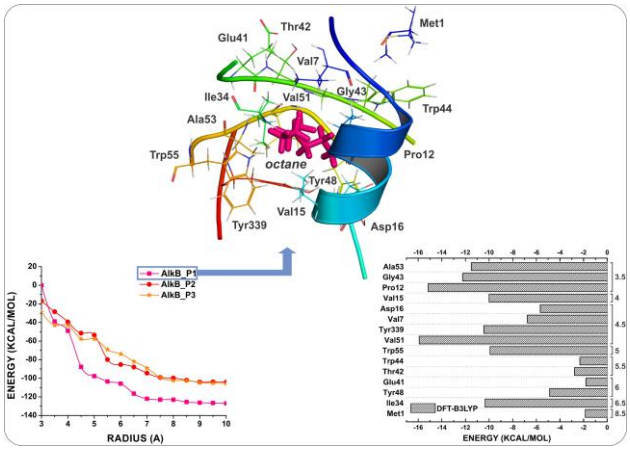
[2] *Departamento de Biofísica e Farmacologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, 59072-970 Natal-RN, Brazil*

[3] *Departamento de Geologia, Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte, 59072-970 Natal-RN, Brazil*

[4] *Departamento de Biologia Celular e Genética, Centro de Biociências, Universidade Federal do Rio Grande do Norte, 59072-970 Natal-RN, Brazil*

**Corresponding author: email: bio.brunogomes@gmail.com*

Graphical abstract



Highlights

- Using protein amino acid sequence of alkane hydroxylase (AlkB) from *Pseudomonas putida* P1 through an *in silico* 3-D homology modelling, docking and MFCC analysis, we characterize and evaluate its interaction energy with the octane, octanol and 1-octyne molecules.
- We detected the preferred position of the octane molecule in Cytosolic Binding Pocket, namely the AlkB_P1 one.
- Our data suggest that Ala53, Trp55, Val15 and Tyr339 of AlkB enzyme are involved with the octane uptake/binding, octanol binding/exit, and the 1-octyne uptake/binding molecule from the alkanotrophic bacteria *P. putida* P1.

Many species of bacteria involved in degradation of n-alkanes have an important constitutional metabolic enzyme, the alkane hydroxylase called AlkB, specialized in the conversion of hydrocarbons molecules that can be used as carbon and/or energy source. This enzyme plays an important role in the microbial degradation of oil, chlorinated hydrocarbons, fuel additives, and many other compounds. A number of these enzymes has been biochemically characterized in detail because the potential of alkane hydroxylases to catalyse high added-value reactions is widely recognized. Nevertheless, the industrial and process bioremediation application of them is restricted, owing to their complex biochemistry, challenging process requirements, and the limited number of their three-dimensional structures. Furthermore, AlkB has great potential as biocatalysts for selective transformation of a wide range of chemically inert unreactive alkanes into reactive chemical precursors that can be used as tools for bioremediation and bioprocesses. Aiming to understand the possible ways the AlkB enzyme *Pseudomonas putida* P1 interacts with octane, octanol and 1-octyne, we consider its suitable biochemical structure taking into account a 3-D homology modelling. Besides, by using a quantum chemistry computational model based on the density functional theory (DFT), we determine possible protein-substrate interaction regions measured by means of its binding energy simulated throughout the Molecular Fractionation with Conjugated Caps (MFCC) approach.

Keywords: alkane hydroxylase, *Pseudomonas putida* P1, 3-D homology modelling, binding energy, quantum chemistry approach.

1. Introduction

Acyclic saturated hydrocarbons, i.e., alkanes, consisting of hydrogen and carbon atoms arranged in a tree structure, are responsible, depending on their geographical origin, to a variation of 20 to 50% of the chemical constitution of the oil. They have the general chemical formula C_nH_{2n+2}. Octane and hexadecane are the best know alkanes, being the most widely used as carbon sources by microorganisms [1]; [2]. They are chemically inert with little biological activity, being viewed as molecular trees upon which can be hung the more active/reactive functional groups of biological molecules. They should

be activated in the presence of oxygen before being metabolized, leading to their two main commercial sources, petroleum (crude oil) and natural gas, by means of the terminal methyl groups' oxidation. The product obtained is oxidized by dehydrogenases to fatty acids, which are metabolized through β-oxidation [3].

Two types of desaturases were identified: a soluble and a membrane one, which are bounded to each other, whose specific catalytic are successfully redesigned in the selection of substrates and catalysis [4]. The enzyme alkane hydroxylase EC 1.14.15.3 (AlkB for short) is a the diiron desaturase whose monomeric form consists of 402 amino acids that is fully anchored in the inner membrane of bacteria alkanotrophics. AlkB protein catalyzes the hydroxylation of n-alkane molecules, like the octane, in alkanol by following the reaction *n-octane + 2 reduced rubredoxin + O_2 + $2H^+$ = 1-octanol + 2 oxidized rubredoxin + H_2O* [5]; [6]; [7].

The use of bacteria that possess the AlkB enzyme system in petroleum bioremediation approaches, as well as in bioprocesses for the synthesis of industrial compounds, are considered two of the most important biotechnological applications of these bacteria [8]. The biotechnological application occurs due to the transformation of the hydrogen and carbon tree structure of the alkanes from the hydrocarbon-contaminated samples. Unfortunately, the rational modification of the AlkB protein for biotechnological use has been precluded due to the lack of detailed information about which residues lining the substrate/product at the enzyme.

There are few techniques that support molecular prediction due to the complexity caused by the number of interactions between them. Among them, computer simulation based on quantum chemistry model has been proved to be a most relevant one [9]; [10]; [11]; [12]. The absence of information on the initial interactions of a substrate to the protein complex makes this technique strategically important in the process of building a model for transport and catalysis. Predicting the position of a substrate to an enzyme through simulations and computational inference based on a quantum chemistry approach is nowadays a promising step for basic and applied research [13]; [14]; [15].

Therefore, as an alternative way, we intend in this work to perform a quantum chemistry calculation of the interaction of the AlkB enzyme *Pseudomonas putida* (*P. putida* for short) P1 with the octane family, that consists mostly of organic compounds obtained from the fractional distillation of petroleum. Among the integral membrane proteins that catalyse the hydroxylation of alkanes, the *P. putida* P1 metabolizes only octane in octanol, simplifying our computer simulation approach because there is no need to consider the high conformational variability of those enzymes capable of accepting various ligands, as is by instance the case of the *P. putida* GPo1, that catalyses the hydroxylation of medium-chain alkanes (C3–C12) [16]. As there is no three-dimensional (3-D) data found in the protein database (PDB) of the alkane hydroxylase, and considering the potentialities of the energy and environmental industries of this enzyme, we firstly perform an in silico analysis of the protein amino acids sequences from alkanotrophics microorganisms using the GenBank database. The search for related protein structures is important for the selection of templates in a model building and evaluation during 3-D structure prediction by homology modelling method. Afterwards, using docking simulations in the newly created AlkB homology model, we identify which residues are the most relevant in a ligand-binding site interaction.

A detailed picture at the quantum level of the binding of the AlkB enzyme with octane molecule (AlkB-octane complex), besides be more accurate, allows a more correct understanding of the role of the receptor amino acid residues and the ligand atoms on the binding pocket features [9]; [15]. Our quantum chemistry model is based on the density functional theory (DFT) calculations within the molecular fractionation with conjugated caps (MFCC) approach to obtain a consistent energetic profile with the most important individual contributions of this complex.

This paper is organized as follows: in section 2, we present the methodology employed here, which is based on several protocols to characterize the alkane hydroxylase protein - AlkB of *P. putida* P1, including some quantum chemistry calculation. Section 3 is devoted to the discussion of the main results of this paper, particularly those related to the binding energies of the AlkB/octane complex. The conclusions of this work are presented in Section 4.

2. Materials and methods

The methodology employed here consists of several model and computer protocols inter-related with each other described below:

2.1 Protein sequence for 3-D homology modelling

Complete sequence of the alkane hydroxylase protein - AlkB of *P. putida* P1 (GI:5531405) were selected from the NCBI GenBank.

2.2 The 3-D Homology modelling

The sequence AlkB protein were modeled by the Swiss-Model server using 1cpt:A atomic data as template. The 1cpt:A is a cytochrome P450terp protein from *Pseudomonas sp*., characterized by a 2.3 Å resolution obtained from a X-ray diffraction technique, whose R-factor is 18.9%, as it can be found at the internet site http://www.ebi.ac.uk/pdbe/entry/pdb/1cpt.

2.3 Visualization and validation of 3-D protein models

All 3-D protein models obtained were visualized and validated with the aid of the Structure Assessment tool of the Swiss-Model server. They were carried out by the fitting between the model and the template 1cpt:A via the software DeepView v3.7, SP5 [17]. For protein visualization we utilized PyMOL v0.99 [18] and the Molegro Virtual Docker (MVD) 2008 [19] programs. We consider the template 1cpt:A for model, whose Procheck parameters were accounted for comparison and validation of the generated models.

The generated 3-D models, by using the fitting method, allowed the degree of structural conservation in relation to the template 1cpt:A, revealing the similarities of the atomic spatial 3-D of the same proteins with low identity in its primary structure. As a consequence, the self-degree of conservation between the enzymes of the same function in different organisms becomes evident, thus complementing the validation of the proposed model.

2.4 Subtrate docking

Docking, by means of the Molegro Virtual Docker (MVD) 2008 software, recently introduced and already with a grown reputation among medicinal chemists, was performed on a 3-D model protein AlkB of *P. putida* P1. It was used together with octane (PubChem CID: 356), octanol (PubChem CID: 957) and 1-octyne (PubChem CID: 12370) from the Structure-Data File (SDF) retrieved from the internet site https://pubchem.ncbi.nlm.nih.gov/. Bench mark results of MVD 2008 software provide very accurate predictions of ligand binding modes (87.0%) compared to other docking software such as Glide (81.8%), GOLD (78.2%), Surflex (75.3%), and FlexX2 (57.9%) [19]. MVD 2008 is based on a differential evolution algorithm called MolDock considering a MolDock Score energy, E_{score} defined by:

$$\boldsymbol{E_{score} = E_{inter} + E_{intra,}} \quad (1)$$

where E_{inter} is the ligand-receptor interaction energy, and E_{intra} is the internal energy of the ligand. Here E_{inter} is given by:

$$\boldsymbol{E_{inter} = \sum_i . \sum_j . \left[E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{(2r_{ij})^2} \right].} \quad (2)$$

In the above equation the summation in *i*,*j* means all ligants and proteins, respectively. Also, the E_{PLP} (r_{ij}) term is a "piecewise linear potential" [20]; [21] characterized by two different parameters. One of them estimates the steric term between the atoms (van der Waals interaction), while the other is responsible for the hydrogen bonds, both terms accounting for the electrostatic interactions between charged atoms [19]. E_{intra} is calculated according to:

$$\boldsymbol{E_{intra} = \sum_i . \sum_j . \left[E_{PLP}(r_{ij}) \right] + \sum_{flexible\ bonds} . A[1 - \cos(m\theta - \theta_0)] + E_{clash,}} \quad (3)$$

where the second summation considers all flexible bonds. Besides, the first term in eq. (3) calculates all energies involving pairs of atoms of the ligand, except those associated with two bonds. The second one represents the torsional energy, where θ is the torsional angle of the bond. The average of the torsional energy bond contributions is considered for the case where several torsions can be determined. The last term, E_{clash}, assigns a penalty of 1000 kcal/mol if the distance between two heavy atoms (more than two bonds apart) is smaller than 2.0 Å, ignoring unfeasible ligand conformations [19]. The molecular docking was performed for all the constituents with the predicted cavities of the substrates. The MolDock score (GRID) function was used with a grid resolution of 0.35 Å and radius of 45 Å with respect to the origin of the respective cavities. The "MolDock SE" algorithm searching for 30 runs, using a maximum of 1500 iterations with a total population size of 50 was applied. The energy threshold used for the minimized final orientation was 100. The simplex evaluation with 300 maximum steps of neighbor distance factor 1 was completed. Initially the whole area includes docking (grid) protein. Later, based on these results, dockings were performed with the grid set to the cavity of greater volume calculated by the software.

Docking protocols suffer from a limited reliability of the prediction of the interaction mode for arbitrary protein−ligand pairs, partly due to the inability to account for protein flexibility and entropic effects in an accurate manner. It is also reported that single rigid receptor dockings predict incorrect binding pose for 50–70% of all ligands [22].

Considering the importance of protein flexibility on ligand binding, the traditional docking routines account for moderate protein flexibility either by allowing partial steric clashes, or by considering independent side-chain rotations [23]. Accordingly, the Molegro Virtual Docker uses soft scoring functions able to tolerate some overlap between the ligand and the protein, which accounts for a small amount of plasticity of the receptor. Regarding the rotatable bonds in each studied ligand-protein complex, all non-cyclic single bonds were set flexible except for bonds that only rotate hydrogens (e.g., bonds connected to hydroxyl and methyl groups).

2.5 Molecular Fractionation with Conjugate Caps (MFCC)

The non-covalent interaction energies among the amino acid residues of an AlkB protein model and the octane substrate in the AlkB_P1, AlkB_P2 and AlkB_P3 binding sites were performed by adopting the widely used dispersion-corrected Becke three-parameter hybrid functional combined with the Lee–Yang–Parr correlation functional (DFT-B3LYP) [24]. A split-valence triple-zeta non-orthogonal one-particle functions polarized basis set, namely the 6-311G*, was used to build the molecular orbitals and the D3 version of Grimme's dispersion with the original D3 damping function (parameters: S6=1.0000, SR6=1.2610, S8=1.7030), including the London dispersion effects in DFT calculations [25]. The B3LYP functional was selected based on previously assessed performance for noncovalent interactions, particularly a weak tendency to overestimate the unfavorable energy of the most repulsive interaction energies [26]. When coupled to the dispersion-correcting potentials to correct the long-range interaction by atom-centered potentials, its accuracy for the S66 test set of small noncovalently bound dimers is excellent [27]. The performance for noncovalent interactions is also encouraging after tests on the HSG benchmark set of 21 dimers and trimers which are present in the complex of the inhibitor indinavir and HIV-II protease [28].

Although the solvation energy and entropy change upon ligand binding are important factors in molecular interactions, unlike the proteins, which depict many charged ionic groups, the AlkB of *P. putida* has predominantly non-charged amino acids, namely 87%, 93% and 80% for AlkB_P1, AlkB_P2 and AlkB_P3 input site, respectively. Due to that, we opted for the gas phase study, enough for the introductory energetic characterization of this region with core predominantly non-polar.

The use of quantum mechanics for molecular modelling and binding interaction analysis has become quite popular in recent years due to its high accuracy in estimating relative binding affinities [29]. However, QM calculations in macromolecules are too time consuming, which demand a balance between the computer execution time and the accuracy of the results. In view of this, it was applied the Molecular Fractionation with Conjugated Caps (MFCC) methodology, which is a very useful approach that provides an accurate description of biological systems through quantum simulations without a very high computational cost [10]; [11]; [12]; [30]; [31].

The interaction (binding) energy between the octane molecule "O" and each amino acid residue "R^i", $E(O\text{-}R^i)$, was calculated according to:

$$E(O - R^i) = E(O - C^{i-1}R^i C^{i+1}) - E(C^{i-1}R^i C^{i+1}) - E(O - C^{i-1}C^{i+1}) + E(C^{i-1} - C^{i+1}), \quad (4)$$

where the C^{i-1} (C^{i+1}) represents the three previously (subsequently) neighboring residues linked to the reference R^i, called 'caps'. The term $E(O\text{-}C^{i-1}R^iC^{i+1})$ corresponds to the total energy of the octane associated with the capped residue. $E(C^{i-1}R^iC^{i+1})$ gives the total energy of the capped residue alone, while the third term, $E(O\text{-}C^{i-1}C^{i+1})$, is the total energy of the system formed by the set of caps and O; finally, $E(C^{i-1}\text{-}C^{i+1})$ is the total energy of the system formed by the isolated caps.

The energy contribution of most important amino acid residues involved in the AlkB-octane complex were plotted in BIRD panels – an acronym for Binding site, Interaction energy and Residues Domain. The BIRD panel depicts, concisely: (i) the interaction energy (in kcal.mol-1) of the residue with the octane, depicted by the horizontal bars, from which one can assess the importance of each interaction/residue in the complex formation; and (ii) the binding interface radius to which each interaction belongs, at the righ side.

3. Results and Discussion

In this section we present the main results of the paper, whose features are discussed below:

3.1 AlkB 3-D homology modelling

Taking into account the amino acid sequences of the AlkB protein from *P. putida* P1, three-dimensional model were obtained by homology modelling using the crystallographic data of 1cpt:A protein from *Pseudomonas sp.* as template (see Fig. 1).

The protein model AlkB was validated by the Swiss-Model server package through a comparative analysis of the Ramachandran-Prochek software for each AlkB protein model generated (see Fig. 2). The data shown by the Ramachandran graph reveal the identity/similarity of primary, secondary and tertiary structures between the 1cpt-A template and the alkane hydroxylase (AlkB) of the *Pseudomonas putida* P1. The 86.3% of AlkB protein residues under study are in the most favorable region, meaning a greater probability that the residues are stereochemically correct in their respective secondary structures. The percentage of amino acids obtained based on the comparative analysis generated by the PROCHEK software allows us to consider that the template used (1cpt-A) is suitable for homology modeling, leading to a reliable 3-D model. According to the parameters related to the main chain adopted as reference by the software, the model was considered within normality, assuming an average resolution of 2.5 Å. There was no indication of incorrect positioning of residues in the model (see Fig. SM1 in the Supplementary Material).

The overall comparison by fitting reveals high structural similarity between AlkB modelling and 1cpt:A protein template, despite the amino-acid sequence low identity. In the literature we found other examples with <10% identity of amino acid sequence [32]; [33]; [34]; [35]. These facts could be explained by the evolutive conservation of the biochemical properties [36], [37]; [38]. After the evaluation of the structural quality of modeled proteins and fitting analysis, we choose AlkB as the N-terminal region due to its proximity to the catalytic site. These regions include 100 amino acids, approximately

arranged in four alpha-helix, three beta strand and random coil structures. We named this region as Cytosolic Binding Pocket (CBP) region (see Figs. 1 and 2).

Interestingly, the CBP region topologically is located in the bacterial cytoplasm, and exhibits the same chemical-physics characteristics of cytosolic 1cpt:A protein, used as template.

3.2 AlkB Docking

To investigate the interaction of n-alkanes with the alkane hydroxylase 3-D model of *P. putida* P1, docking simulations were carried out on CBP, with the grid adjusted to the cavity of greater volume (1951.88 $Å^3$). After thirty molecular simulations, the octane substrate exhibits the better interaction energy (approximately -70 kcal/mol). According to the octane molecules linked to the CBP region, four different preferred positions were verified. Among them, three positions were chosen by considering statistical criteria that take into account the interaction energy between the template and the substrate, namely: AlkB_P1, AlkB_P2 and AlkB_P3. The position AlkB_P4 was dismissed because it exhibits fewer octane links (see Fig. 3).

To deepen the analysis of CBP we performed also docking simulation with octanol molecules (product from octane hydroxilation), as well as with 1-octyne molecule, which is known to inhibit the enzyme AlkB [2]; [39]; [40]. Our results, from docking simulation by MVD 2008, allow the identification in the CBP of 10 amino acids involved in the interaction with the substrate (octane), inhibitor (1-octyne) and product (octanol) molecules, separately. Among them, four amino acids residues are shared by these molecules: Ala53, with energies -2.87, -5.15, and -2.75 kcal/mol; Trp55**,** with energies -4.84, -14.61 and -4.70 kcal/mol; Val15**,** with energies -3.75, -5.04 and -9.84 kcal/mol; and Tyr339**,** with energies -5.35, -14.01 and -12.31 kcal/mol for the octane, 1-octyne and octanol, respectively (see Table 1 and Fig. 4).

The total energy found on the interaction between these 10 aminoacids and the octane (-67,146 kcal/mol), the octanol (-55,654 kcal/mol), and the 1-octyne (-67,0369 kcal/mol) at CBP region are highly suitable for binding and stability of the molecular arrangement and the enzime. These results are the first report with AlkB enzyme *P. putida* P1 from: (i) amino acid interaction with octane, octanol and 1-octyne molecules, (ii) interaction energy data and (iii) molecular positioning at CBP region.

These finding suggest that this region has stereochemical compatibility, not only for early 1-octyne molecule interaction, but also for octanol exit and octane entry via cytosol, the latter being an alternative possibility.

On the other hand, we verified the proximity of the molecules under study with the histidine motif of AlkB [2]; [40]. According to van Beilen *et al*. (van Beilen & Funhoff, [40], AlkB enzyme has four histidine motifs involved to bind the diiron cluster responsible for oxidation of the substrate. Location of the octanol molecule on CBP region was obtained by docking, and reveals that its hydroxyl terminal is oriented to histidine residues. One of them, the His318**,** has energy -4.62 kcal/mol, and is distant 5.06 Å of the octanol molecule, while the others three (His312, His315 and His316)

belonging to the histidine motif D, depict weakly interaction with the product (see Table 1 and Fig. 4).

Recently Alonso *et al*. [39] showed a 3-D computer model of AlkB protein from *P. putida* GPo1, a genetically diferent alkanotrophic bacteria of *P. putida* P1. These authors mapped the location of the active site of AlkB by 1-octyne inhibitor, missing the ultimate amino acid interaction with AlkB enzyme, as well as with the octane and the octanol molecules.

Our *in silico* data suggest that this CBP region is a local binding for substrate passage to catalytic site. They also expand not only the structural model of AlkB for *P. putida* P1 proposed by van Beilen *et al*. [2]; [40], but also gives an insight of the CBP region regarding the specific amino acid interaction with the octane, the octanol, and the 1-octyne molecules.

3.3 MFCC analysis

The MFCC approach was applied in the selected areas of alkanes obtained for docking in AlkB enzyme, namely AlkB_P1, AlkB_P2 and AlkB_P3. The convergence study of the total interaction energy as a function of the ligand binding pocket radius r was performed in order to put a limit in the number of amino-acid residues to be analyzed without missing important interactions. In doing that, we have added the individual interaction energy of the amino-acid residues within imaginary spheres with pocket radius r centered at the ligand. Considering the ligand binding pocket radius $r = n/2$ (where n = 1, 2, 3, 4...), we achieved the converged binding pocket radius when the energy variation in the subsequent radius is smaller than 10% [9]; [41]. According to our simulations, this behavior was observed for distances in the 9–10 Å range, encompassing a total of 90, 105 and 103 amino acids residues for AlkB_P1, AlkB_P2 and AlkB_P3, respectively (see Fig. 5).

Therefore, we believe that the residues located at distances larger than 9 Å do not contribute significantly to the AlkB-octane binding interaction. For residues inside this closed binding pocket radius, the total interaction energy follows the order AlkB_P1 (-124.82 kcal/mol) > AlkB_P2 (-104.11 kcal/mol) ≈ AlkB_P3 (-104.56 kcal/mol).

For AlkB_P1, the presence of eight of the most strongly attractive amino acid residues in the region covering the range 3 to 5 Å, induces an abrupt decrease of -97.67 kcal/mol in the total interaction energy, corresponding to 72% of the total binding energy. These residues form a hydrophobic channel with high affinity binding energy (see Fig. 6), namely: Val51 (-15.92 kcal/mol), Gly43 (-12.24 kcal/mol), Ala53 (-11.51 kcal/mol), Tyr339 (-10.45 kcal/mol), Ile34 (-10.36 kcal/mol) Val15 (-10.01 kcal/mol), Trp55 (-9.93 kcal/mol) and Val7 (-6.77 kcal/mol). Thus, it can be expected that the van der Waals interactions will play an important role in AlkB/octane complex. Indeed, we believe that hydrophobic contributions can be quite large in complexes with receptors containing cavities: the release of high-energy water molecules inside such cavities is essentially responsible for the very high affinities of the ligand-receptor complexes [42]; [43].

Similarly to this work, the London dispersion interactions have an important impact in the structural stability of biomolecular systems [44]; [45], biological membranes [46], supramolecular complexes [45] and interaction energy of ionic liquids [43]. The

residues Pro12 (-15,16 kcal / mol), Asp16 (- 5,67 kcal / mol), Tyr48 (- 4,88 kcal / mol), Thr42 (- 2,76 kcal / mol), Trp44 (- 2,32 kcal / mol), Met1 (- 1,86 kcal / mol), Glu41 (- 1,81 kcal / mol) complete the list of those most relevant to the complexation of octane in AlkB_P1.

For completeness, we also present in the Supplemental Material enclosed, the MFCC interaction energy between the octane and the most relevant AlkB residues in AlkB_P2 and AlkB_P3 at the Cytosolic Binding Pocket (CBP) region (see Fig. SM2 and SM3 in the Supplementary Material), depicting the following residues that bind more significantly: Pro319 (-12,05 kcal / mol), Ile371 (-11,75 kcal / mol), Gln372 (-11,23 kcal / mol), Pro60 (-10,47 kcal / mol), Ala64 (-8,53 kcal / mol), Pro340 (-8,20 kcal / mol), Ala317 (-7.59 kcal / mol), Asn72 (-4.85 kcal / mol), Phe70 (-4.01 kcal / mol), Ile373 (-3.38 kcal / mol), Leu368 (-2.90 kcal / mol), Met345 (-2.83 kcal / mol), His318 (-2.25 kcal / mol), Asp63 (-2.08 kcal / mol) and Met377 (-1.79 kcal / mol) for Octano-AlkB_P2 complex; and Tyr339 (-22.86 kcal / mol), Val15 (-14,25 kcal / mol), Pro12 (-9,62 kcal / mol), Val51 (-9,60 kcal / mol), Lys18 (-6,16 kcal / mol), Asp16 (-5,95 kcal / mol), His316 (-5,93 kcal / mol), Gly341 (-5,59 kcal / mol), Glu79 (-5,00 kcal / mol), Gly43 (-4,86 kcal / mol), Ala53 (-3,71 kcal / mol), Met33 (-1,90 kcal / mol), Met1 (-1,87 kcal / mol), Phe21 (-1,82 kcal / mol) and His315 (-1,80 kcal / mol) for Octano-AlkB_P3 complex.

Figures 6 and 7 reveal a tight temporary binding affinities in AlkB_P1. In fact, we highlight the london dispersion forces between octane and nonpolar amino acids residues, such as Val51, Pro12, Gly43, Ala53, Tyr339, Ile34, Val15, Trp55 and Val7. This fact corroborates the hypothesis that hydrophobic input channels are preferred regions for hydrocarbons in AlkB integral membrane non-heme iron monooxygenase. According to the data obtained by the MFCC method, the best position at CBP site is AlkB_P1.

After the docking simulation by MVD 2008 methods and MFCC analysis of AlkB_P1, we detected eight amino acids, placed on hydrophobic channel, revealing a better interaction with the octane. Noteworthy, by docking simulation using the MVD methods, we observe that four of them, namely Ala53, Trp55, Val15 and Tyr339 also display high interaction energy with two other important molecules, the octanol and the 1-octyne one.

Particularly, Trp55 residue was mutated in *P. putida* (AlkB) and Trp58 in *Alcanivorax borkumensis* (AlkB1) [47]. These residues are very important to the discrimination of n-alkane substrate range, enabling the *P. putida* to use longer n-alkanes than the original strain [40]. These data support the importance of Trp55 residue interacting with the substrate, product and inhibitor molecules at the CBP. Based on these interaction energies and the spatial positioning data, the hypothesis that the CBP can be a transit location of octane for hydroxylation site is highly supported.

It is important to mention here the work of Alonso et al. [39] which provided convincing experimental/computational results supporting the role of protonated Lys18 in the hydroxylation of medium-chain alkanes by *Pseudomonas putida* GPo1 alkane hydroxylase (AlkB). When compared with the wild-type protein, AlkB K18A, it has

lower thermodynamic stability, a less compact fold and an altered secondary structure, suggesting that the loss of enzymatic activity in the K18A variant is due to the loss of structural integrity. However, our study was not performed in the region of the active site of the hydroxylase. Indeed, Alonso and co-workers presented the active site within the transmembrane domain of AlkB by 1-octyne inhibitor. On the other hand, our in silico data suggest a cytosolic binding region, namely AlkB_P1, for substrate (octane) passage to catalytic site, actually complementing the findings of their previous investigation.

4. Conclusions

Summarizing, using protein amino acid sequence of alkane hydroxylase (AlkB), EC 1.14.15.3, from *P. putida* P1 through an *in silico* 3-D homology modelling, docking and MFCC analysis, we characterize and evaluate its interaction energy with the octane, octanol and 1-octyne molecules. Through the 3-D homology modelling it was possible to predict the atomic structural arrangement of AlkB enzyme particularly in the cytoplasmic region. Docking simulation revealed three preferred positions in a defined binding pocket region, besides a more refined assessment using quantum chemistry computational. Furthermore, our data suggest that Ala53, Trp55, Val15 and Tyr339 of AlkB enzyme are involved with the octane uptake/binding, octanol binding/exit, and the 1-octyne uptake/binding molecule from the alkanotrophic bacteria *P. putida* P1.

On the other hand, through an MFCC analysis, we detected the preferential position of octane molecule in the Cytosolic Binding Pocket, AlkB_P1. By docking simulation using the MVD 2008 software, it was observed high interaction energy with two other very important molecules: octanol (product) and the 1-octyne (inhibitor). We detected the preferred position of the octane molecule in BPC, namely the AlkB_P1 one (see Fig. 6), whose data infer a possible transit point (substrate / product / inhibitor molecules). Individually, each substrate molecule (octane) product (octanol) and inhibitor (1-octyne) interact simultaneously with Ala53, Trp55, Val15 and Tyr339 of AlkB enzyme. An alternative possibility to alkane uptake from periplasmic space implies that the entry to hydroxylation may occur in cytosolic environment, giving evidence of a possible cytosolic pathway for alkane molecule AlkB enzyme of *P. putida* P1.

We believe that our detailed characterization of the binding energies between the AlkB enzyme *Pseudomonas putida* P1 and the octane, octanol and 1-octyne structures is a step forward to understand the general processes of interaction in these complex environments.

Funding: This work was supported by the CAPES, CNPq, Projeto Universal/CNPq and (Rede 05 PETROMAR, FINEP/CNPq/PETROBRAS).

References

[1] So, C. M. & Young, L. Y. Isolation and characterization of a sulfate-reducing bacterium that anaerobically degrades alkanes. *Applied and Environmental Microbiology*, 1999, 65, 2969–2976.

[2] van Beilen, J. B., Li, Z., Duetz, W. A., Smits, T. H. M, & Witholt, B. Diversity of alkane hydroxylase system in the environment. *Oil & Gas Science and Technology*, 2003, 58, 427–440.

[3] van Hamme, J. D., Singh, A., & Ward, O. P. Recent advances in petroleum microbiology. *Microbiology and Molecular Biology Reviews*, 2003, 67, 503–549.

[4] Shanklin, J., & Cahoon, E. B. Desaturation and related modifications of fatty acids. *Annual Review of Plant Physiology and Plant Molecular Biology*, 1998, 49, 611–641.

[5] Cardini, G., & Jurtshuk, P. The enzymatic hydroxylation of n-octane by Corynebacterium sp. strain 7E1C. *The Journal of Biological Chemistry*, 1970, 245, 2789–2796.

[6] McKenna, E. J., & Coon, M. J. Enzymatic ω-oxidation. IV. Purification and properties of the ω-hydroxylase of Pseudomonas oleovorans. *The Journal of Biological Chemistry*, 1970, 245, 3882–3889.

[7] Peterson, J. A., Kusunose, M., Kusunose, E., & Coon, M. J. Enzymatic ω-oxidation. II. Function of rubredoxin as the electron carrier in ω-hydroxylation. *The Journal of Biological Chemistry*, 1967, 242, 4334–4340.

[8] Cao, B.; Nagarajan, K., & Loh, K. C. Biodegradation of aromatic compounds: current status and opportunities for biomolecular approaches. *Applied Microbiology and Biotechnology*, 2009, 85, 207–228.

[9] Neto, J. X. L, Fulco, U. L., Albuquerque, E. L., Corso, G., Bezerra, E. M., Caetano, E. W. S., Costa, R. F., & Freire, V. N. A Quantum Biochemistry Investigation for Willardiine Partial Agonism in AMPA Receptors. *Physical Chemistry Chemical Physics*, 2015, 17, 13092–13103.

[10] Rodrigues, C. R. F., Oliveira, J. I. N., Fulco, U. L., Albuquerque, E. L., Moura, R. M., Caetano, E. W. S., & Freire, V. N. Quantum biochemistry study of the T3-785 tropocollagen triple-helical structure. *Chemical Physics Letters*, 2013, 559, 88–93.

[11] Mota, K. B., Neto, J. X. L., Costa, A. H. L., Oliveira, J. I. N., Bezerra, K. S., Albuquerque, E. L., Caetano, E. W. S., Freire, V. N., & Fulco, U. L. A quantum biochemistry model of the interaction between the estrogen receptor and the two antagonists used in breast cancer treatment. *Computational and Theoretical Chemistry*, 2016, 1089, 21–27.

[12] Ouriques, G. S., Vianna, J. F., Neto, J. X. L., Oliveira, J. I. N., Mauriz, P. W., Vasconcelos, M. S., Caetano, E. W. S., Freire, V. N., Albuquerque, E. L., & Fulco, U. L. A quantum chemistry investigation of a potential inhibitory drug against the dengue virus. *RSC Advances*, 2016, 6, 56562–56570.

[13] Xiang, Y., Zhang, D. W., & Zhang, J. Z. Fully quantum mechanical energy optimization for protein-ligand structure. *Journal of Computational Chemistry*, 2004, 25, 1431–1437.

[14] Wu, E. L., Mei, Y., Han, K., & Zhang, J. Z. Quantum and molecular dynamics study for binding of macrocyclic inhibitors to human alpha-thrombin. *Biophysical Journal*, 2007, 92, 4244–4253.

[15] Dantas, D. S., Oliveira, J. I. N., Neto, J. X. L., Costa, R. F., Bezerra, E. M., Freire, V. N., Caetano, E. W. S., Fulco, U. L., & Albuquerque, E. L. Quantum molecular modelling of ibuprofen bound to human serum albumin. *RSC Advances*, 2015, 5, 439–450.

[16] Witholt, B., Smet, M. J., Kingma, J., van Beilen, J. B., Kok, M., Lageveen, R. G. & Eggink, G. Bioconversions of aliphatic compounds by Pseudomonas oleovorans in multiphase bioreiactors: background And economic potential, *Trends in Biotechnology*, 1990, 8, 46-52.

[17] Guex, N. and Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, 1997, 18, 2714-2723.

[18] De Lano, W.L. The PyMOL Molecular Graphics System. Version 0.99rc6 Schrödinger, LLC, *De Lano Scientific*, San Carlos, 2002.

[19] Thomsen, R., & Christensen, M. H. MolDock: a new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, 2006, 49, 3315–3321.

[20] Gehlhaar, D. K., Verkhivker, G., & Rejto, P. A. Fully automated and rapid flexible docking of inhibitors covalently bound to serine proteases. International Conference on Evolutionary Programming. *Springer Berlin Heidelberg*, 1998, 449-461.

[21] Yang, J., & Chen, C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Structure, Function and Genetics*, 2004, 55, 288–304.

[22] Totrov, M., & Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology*, 2008, 18, 178–184.

[23] B-Rao, C., Subramanian, J., & Sharma S. D. Managing protein flexibility in docking and its applications. *Drug Discovery Today*, 2009, 14, 394-400.

[24] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 1993, 98, 5648–5652.

[25] Grimme, S., Antony, J., Ehrlich, S., & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics*, 2010, 132, 154104.

[26] Li, A., Muddana, H. S., & Gilson, M. K. Quantum mechanical calculation of noncovalent interactions: A large-scale evaluation of PMx, DFT, and SAPT approaches. *Journal of chemical theory and computation*, 2014, 10(4), 1563-1575.

[27] Rezáč, J., Riley, K. E., & Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *Journal of Chemical Theory and Computation*, 2011, 7, 2427-2438.

[28] Faver, J. C., Benson, M. L., He, X., Roberts, B. P., Wang, B., Marshall, M. S., Kennedy, M. R., Sherrill, C. D., & Merz, K. M. Jr. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-ligand Complexes. *Journal of Chemical Theory and Computation*, 2011, 7, 790-797.

[29] Gordon, M. S., Fedorov, D. G., Pruitt, S. R., & Slipchenko, L. V. Fragmentation methods: a route to accurate calculations on large systems. *Chem. Rev*, 2012, 112(1), 632-672.

[30] Zhang, D. W., & Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *The Journal of Chemical Physics*, 2003, 119, 3599–3605.

[31] Costa, R. F., Freire, V. N., Bezerra, E. M., Cavada, B. S., Caetano, E. W. S., Filho, J. L. L., & Albuquerque, E. L. Explaining Statin Inhibition Effectiveness of HMG-CoA Reductase by Quantum Biochemistry Computations. *Physical Chemistry Chemical Physics*, 2012, 14, 1389–1398.

[32] Brenner,S.E. et al. Understanding protein structure: using scop for fold interpretation. *Methods Enzymol*., 1996, 266, 635–643.

[33] Holm,L. and Sander,C. Mapping the protein universe. *Science*, 1996, 273, 595–603.

[34] Hubbard,T.J.P. et al. SCOP: a structural classification of proteins database. *Nucleic Acids Res*., 1997, 25, 236–239.

[35] Valencia,A. et al. GTPase Domains of Ras p21 Oncogene Protein and Elongation Factor Tu: Analysis of Three-Dimensional Structures, Sequence Families, and Functional Sites. *Proc. Natl Acad. Sci. USA*, 1991, 88, 5443–5447.

[36] Brejc,K. et al. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, 2001, 411, 269–276.

[37] Celie,P.H.N. et al. Crystal structure of acetylcholine-binding protein from Bulinus truncatus reveals the conserved structural scaffold and sites of variation in nicotinic acetylcholine receptors. *J. Biol. Chem*., 2005, 280, 26457–26466.

[38] Krissinel, E. On the relationship between sequence and structure similarities in proteomics. *Bioinformatics*, 2007, 23, 6, 717–723.

[39] Alonso, H., Kleifeld, O., Yeheskel, A., Ong, P. C., Liu, Y. C., Stok, J. E., De Voss, J. J., & Roujeinikova, A. Structural and mechanistic insight into alkane hydroxylation by *Pseudomonas putida* AlkB. *Biochemical Journal*, 2014, 460, 283–293.

[40] van Beilen, J. B., & Funhoff, E. G. Expanding the alkane oxigenase toolbox: new enzymes and applications. *Current Opinion in Biotechnology*, 2005, 16, 308–314.

[41] Zahn, S. & Kirchner, B. Validation of Dispersion-Corrected Density Functional Theory Approaches for Ionic Liquid Systems. *The Journal of Physical Chemistry A*, 2008, 112, 8430–8435.

[42] Schneider, H. J. Dispersive Interactions in Solution Complexes. *Accounts of Chemical Research*, 2015, 48, 1815–1822.

[43] Biedermann, F., Nau, W. M., & Schneider, H. The Hydrophobic Effect Revisited—Studies with Supramolecular Complexes Imply High-Energy Water as a Noncovalent Driving Force. *Angewandte Chemie International Edition*, 2014, 53, 11158–11171.

[44] Kolář, M., Kubař, T., & Hobza, P. On the Role of London Dispersion Forces in Biomolecular Structure Determination. *The Journal of Physical Chemistry B*, 2011, 115, 8038–8046.

[45] Steinmann, S. N., & Corminboeuf, C. Comprehensive Benchmarking of a Density-Dependent Dispersion Correction. *Journal of Chemical Theory and Computation*, 2011, 7, 3567–3577.

[46] Wagner, J. P., & Schreiner, P. R. Nature Utilizes Unusual High London Dispersion Interactions for Compact Membranes Composed of Molecular. *Ladders Journal of Chemical Theory and Computation*, 2014, 10, 1353–1358.

[47] van Beilen, J. B., Smits, T. H. M., Roos, F. F., Brunner, T., Balada, S. B., Röthlisberger, M., et al. Identification of an amino acid position that determines the substrate range of integral membrane alkane hydroxylases. *J. Bacteriol*, 2005, 187, 85–91.

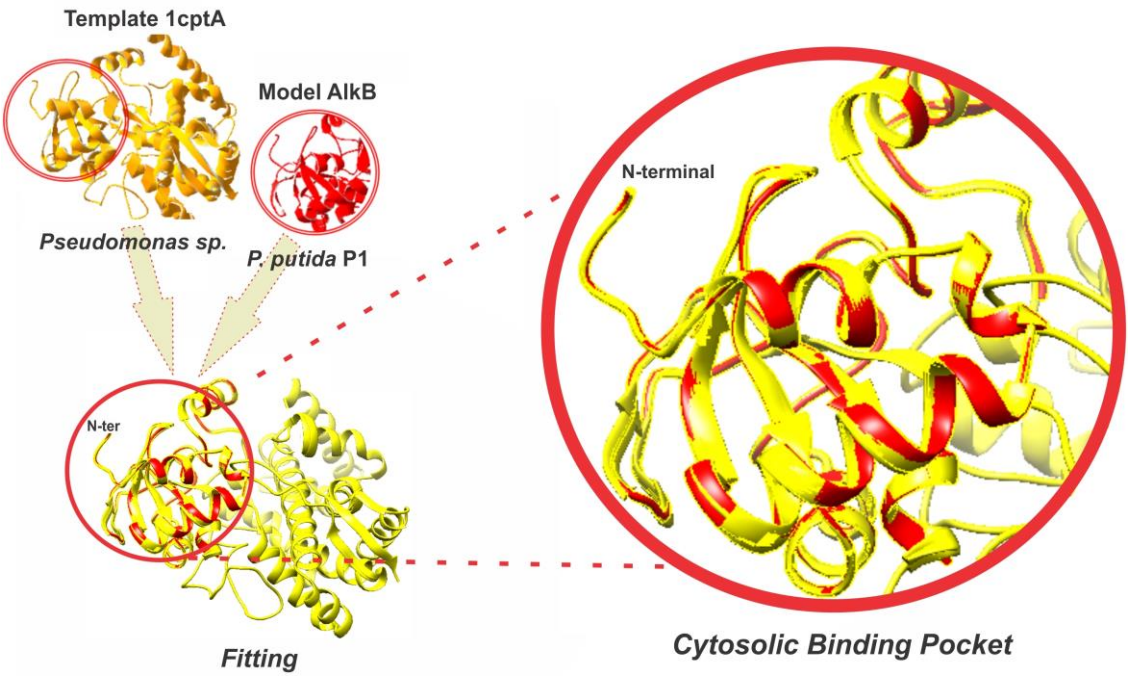
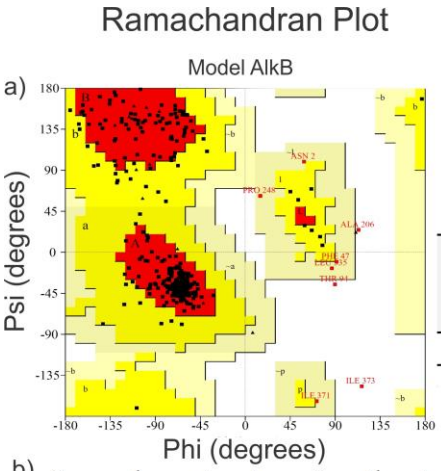


Figure 1: 3-D models generated by the Swiss-Model server using PDB 1cpt data. High structural similarity was observed after fitting on the Cytosolic Binding Pocket (CBP) of proteins.



Comparative statistical analysis of Ramachandran plots generated for validation of protein

Model	Residues in most favoured regions [A,B,L]	Residues additional allowed regions [a,b,l,p]	Residues in generously allowed regions [~a,~b,~l,~p]	Residues disallowed regions	Number of non-glycine and non-proline residues	Number of end-residues (excl. Gly and Pro)	Number of glycine residues (shown as triangles)	Number of proline residues	Total number of residues
AlkB	302 (86,3%)	41 (11,7%)	4 (1,1%)	3 (0,9%)	350 (100%)	2	29	21	402

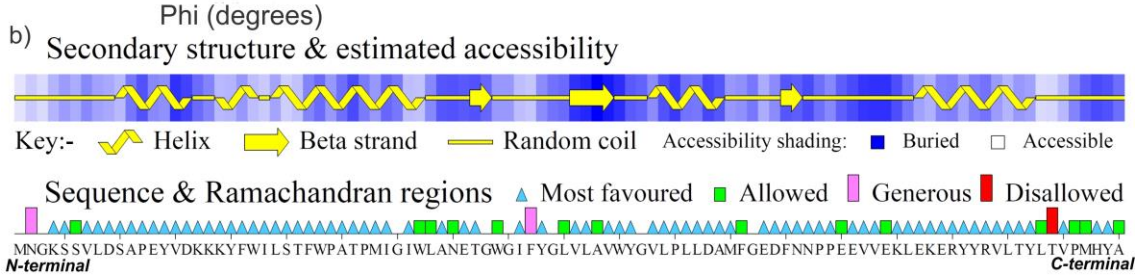


Figure 2: a) Graphs obtained by using the Ramachandran-Prochek software for each AlkB protein model generated. To get a model of good quality (above 90%), we have considered the analysis of 118 structures with a resolution of at least 2.0 Angstroms and R-factor no greater than 20. b) Evaluation of the structural quality of modeled protein generated by the Ramachandran-Prochek software.

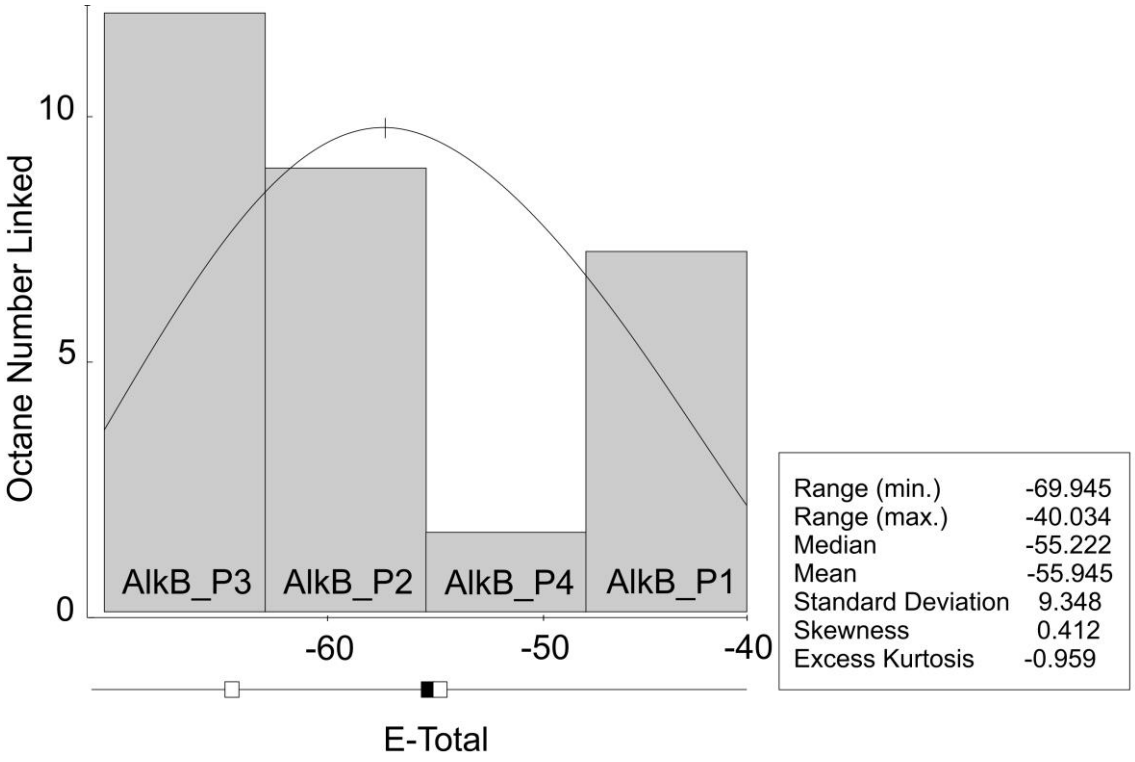


Figure 3: The octane linked to the distribution graph CBP according to the interaction energies**.** From the thirty docking simulations, three positions were chosen in the sequence AlkB_P3 < AlkB_P2 < AlkB_P1.

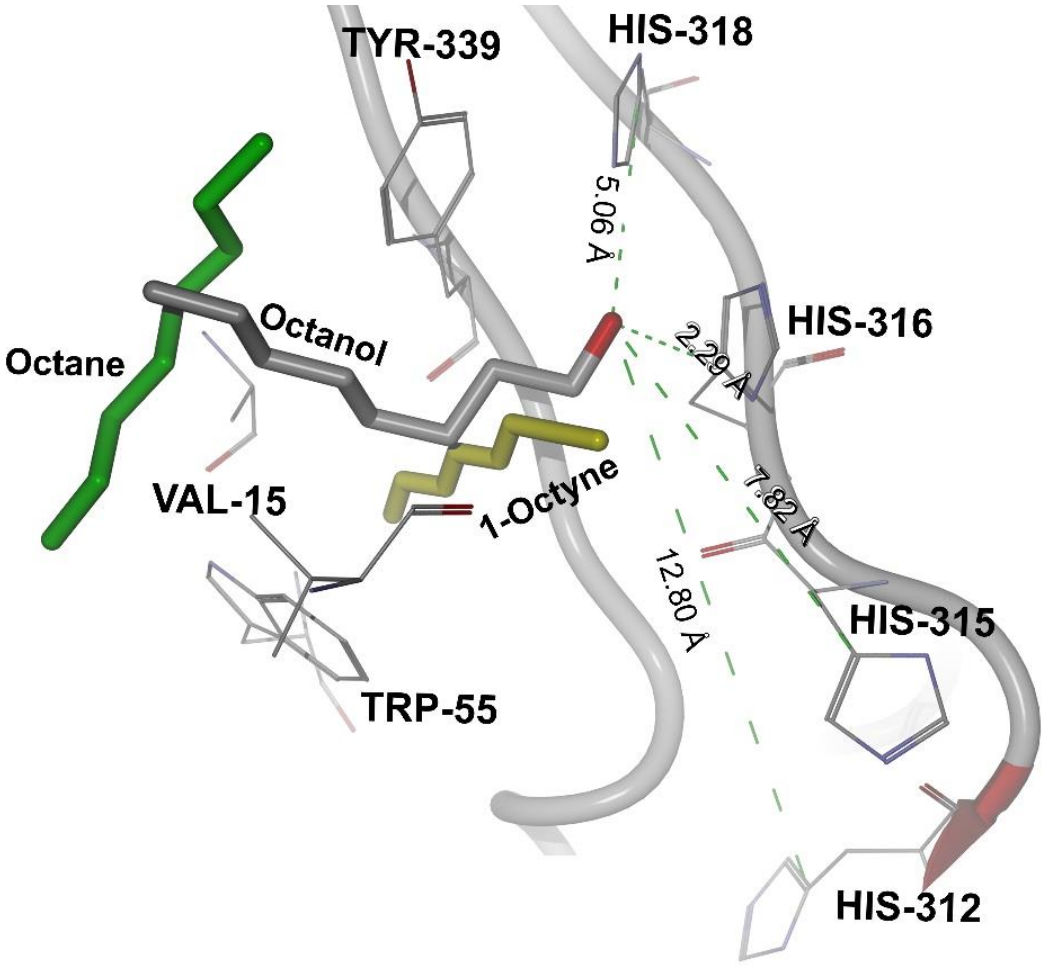


Figure 4: Spatial arrangement of the substrate molecules (octane), product (octanol) and inhibitor (1-octyne) obtained by docking in CBP. Here, the amino acids that interact with the three molecules are highlighted. The distance of the histidine residues in relation to the hydroxy terminal portion of the molecule octanol are shown by dashed lines.

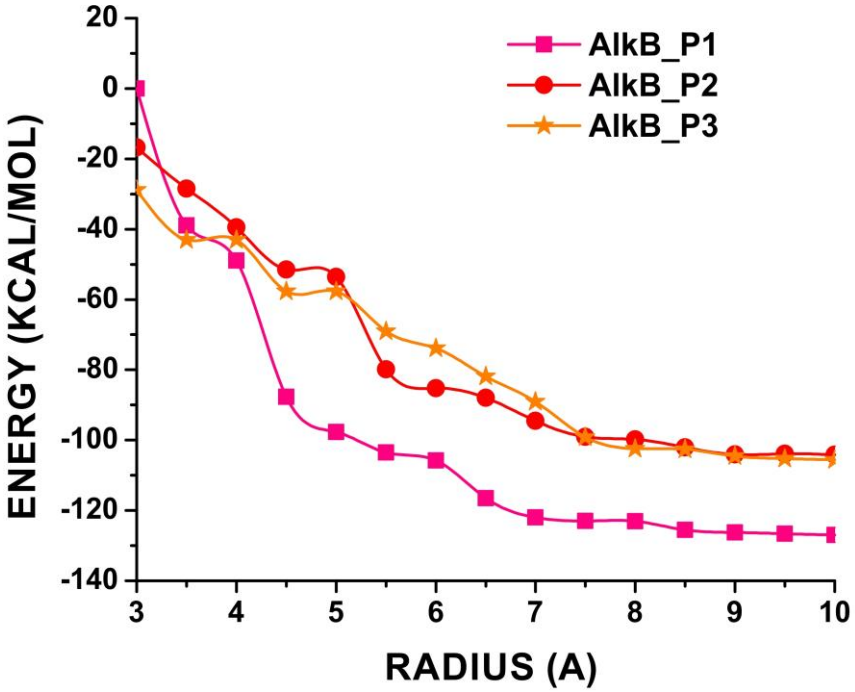


Figure 5: Variation of the interaction energy as a function of the distance for octane. The binding pocket radius are given by the distance between the octane centroid and the most distant AlkB residue binding ligand.

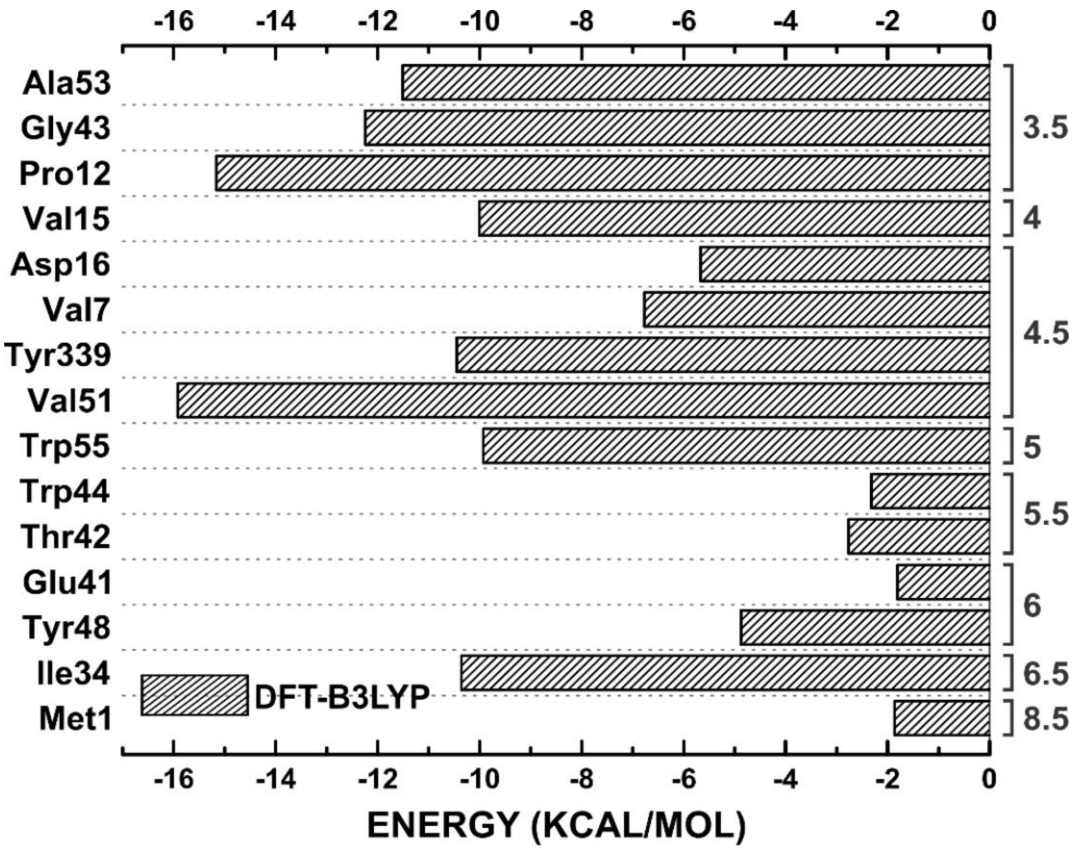


Figure 6: Binding site, interaction energy and residues domain (BIRD) panel showing the MFCC interaction energy between the octane and the most relevant AlkB residues in AlkB_P1 at CBP.

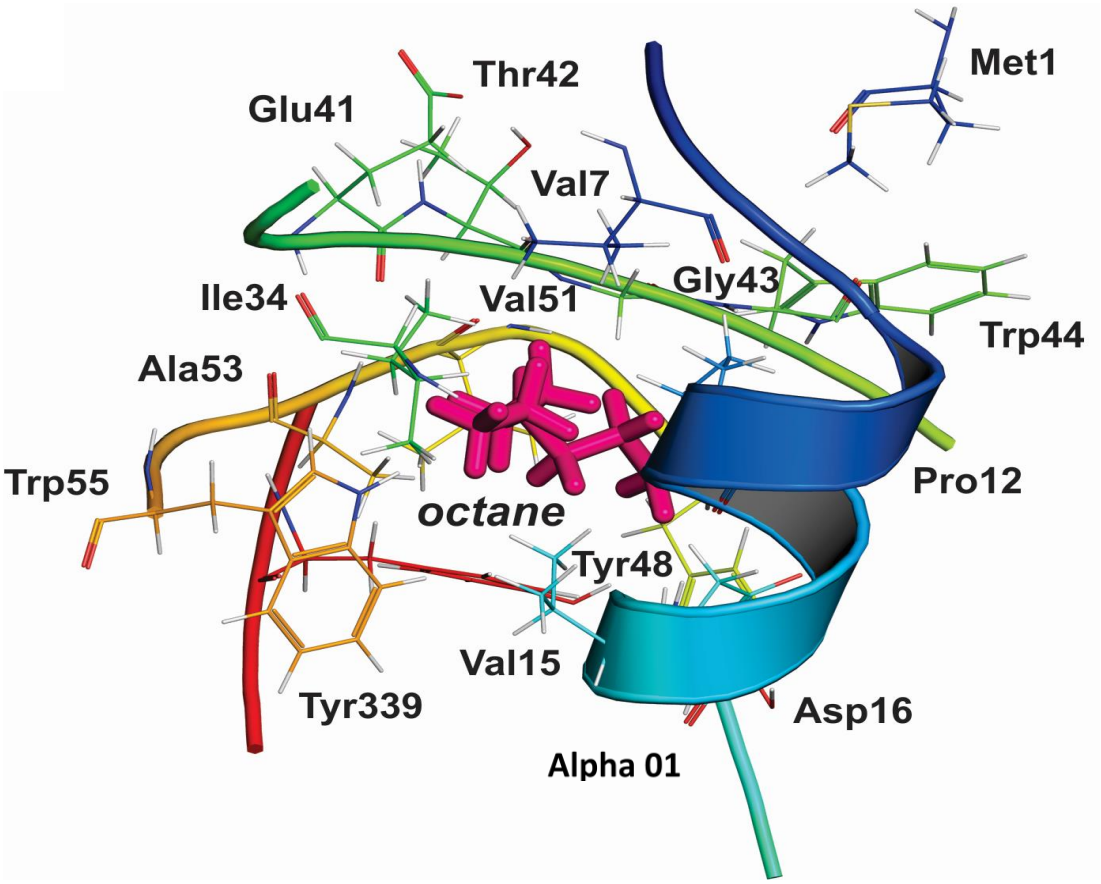


Figure 7: Spatial arrangement of the amino acid residues with largest contribution to the binding energy at the AlkB_P1 input site.

Table 1: Interaction energy of the substrate (octane), inhibitor (1-octyne) and product (octanol) molecules in the CBP

Interaction energy in CBP (kcal/mol)			
AA	**DOCKING MVD**		
	OCTANE P1	**1-OCTYNE**	**OCTANOL**
Pro12	-11.9936	NC	-4.6505
Gly43	-6.5404	NC	-2.4523
Val7	-5.5790	NC	-0.4849
Ala53 #	-2.8719	-5.147	-2.7508
Trp55 #	-4.8405	-14.613	-4.6986
Val15 #	-3.7483	-5.040	-9.8402
Trp44	-2.8491	NC	-0.4724
Tyr339 #	-5.3507	-14.012	-12.3077
Val51	-5.2302	NC	-1.9483
Asp16	-5.1778	NC	-10.2927
E-Total*	**-67.196**	**-67.0369**	**-55.654**

(AA) Amino Acid.
(NC) Not Compared (low significance or no interaction due to the spatial positioning of the molecule).
(*) Total interaction energy of all amino acids that interact with the molecules, including residues not shown here.