Computational and Structural Biotechnology Journal 17 (2019) 757–760

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Communications

Prediction of Ligand Transport along Hydrophobic Enzyme Nanochannels

Diego E. Escalante [a], Alptekin Aksan [a,b,*]

[a] Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55455, United States.
[b] BioTechnology Institute, University of Minnesota, St. Paul, MN 55108, United States.

ARTICLE INFO

Article history:
Received 20 February 2019
Received in revised form 8 May 2019
Accepted 1 June 2019
Available online 11 June 2019

ABSTRACT

Buried active sites of enzymes are connected to the bulk solvent through a network of hydrophobic channels. We developed a discretized model that can accurately predict ligand transport along hydrophobic channels up to six orders of magnitude faster than any other existing method. The non-dimensional nature of the model makes it applicable to *any* hydrophobic channel/ligand combination.

1. Introduction

Nanochannels inside enzymes are responsible for controlling the bidirectional transport of small molecules (ligands) between the buried active sites and the cell's cytoplasm [1,2]. These nanochannels may also serve more sophisticated functions such as preventing competing side reactions [2], protecting the enzyme from toxic or unstable intermediates [3], and selecting for substrates [2]. During the last decade, several software tools were developed for the identification and characterization of nanochannels [4]. However, none of these software were designed to model the transport of ligands through nanochannels and enable rapid determination of whether a ligand is capable of reaching the active site. The lack of such a modeling tool necessitates screening and identification of novel substrates using experimental [5] and computational [6–8] approaches that are expensive and time-consuming.

In this communication, we describe a coarse-grained model for prediction of ligand transport inside hydrophobic enzyme nanochannels that is faster than the all-atom [8] and steered molecular dynamics [7] alternatives. To reduce the excessive computational requirement for calculating all pairwise interaction potentials, we perform a simple discretization (slicing) procedure with which a hydrophobic channel inside an enzyme is represented as a sequence of building blocks as shown in Fig. 1a. Each building block is defined by three parameters (Fig. S1) to describe its geometry and physicochemical characteristics: i) the entrance radius (r_i); ii) the midpoint radius (r_o); and iii) the intermolecular nonbonded interaction strength (ε). The nonbonded interaction strength of the building block, ε_C, is defined in terms of the Lennard-Jones potential. Similarly, the ligand is modeled as a sphere of uniform hydrophobicity represented by the nonbonded interaction strength, ε_L. We nondimensionalized the building block geometric parameters (e.g. r_o/r_i); and the nonbonded strengths of the building block, and the ligand with respect to the potential well of a SPC/E water molecule ($\varepsilon_C/\varepsilon_W$, and $\varepsilon_L/\varepsilon_W$, respectively). In addition, the volume fraction of the building blocks inaccessible to water molecules (i.e. the excluded volume, V_O/V_T) was found to be a critical parameter in modeling the transport of ligands. The dimensional analysis allowed the generation of a unified set of topologies that can describe *any* given hydrophobic channel section/ligand combination. A more detailed explanation of the building block parameters and the dimensional analysis can be found in the methods section. Excluded volume values for each building block are provided in Table S1.

The non-linear regression in Fig. 1b shows the correlation between the nondimensional Gibbs free energy of transport ($\Delta G^* = \Delta G/k_BT$) and the dimensionless parameters that characterizes the contributions of geometry and hydrophobicity of the system, as well as exclusion volume effects inside the building blocks.

$$\Delta G^* = \log\left[\left(\frac{\varepsilon_C}{\varepsilon_W}\right)^4\left(\frac{\varepsilon_C}{\varepsilon_L}\right)^{1.5}\left(\frac{r_o}{r_i}\right)^3\left(\frac{V_O}{V_T}\right)\right] = 0 \quad (1)$$

Eq. (1) defines the cutoff between unsuccessful, and successful transport across the building blocks (Shown in Fig. 1b as the shaded, and unshaded areas, respectively). The cutoff was based on observations made during our building block model development [9], where ligands did not traverse the full length of the building block at certain geometry/hydrophobicity combinations. Details on the derivation of Eq. (1) are provided in the methods section.

To test the accuracy of our ligand transport model, we selected the enzyme naphthalene 1,2-dioxygenase (NDO), natively expressed in *Pseudomonas putida* NCIB 9816-4, as the model enzyme. It has been shown theoretically [10], and experimentally [11] that substrate

* Corresponding author at: Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55455, United States.
E-mail address: aaksan@umn.edu (A. Aksan).

https://doi.org/10.1016/j.csbj.2019.06.001

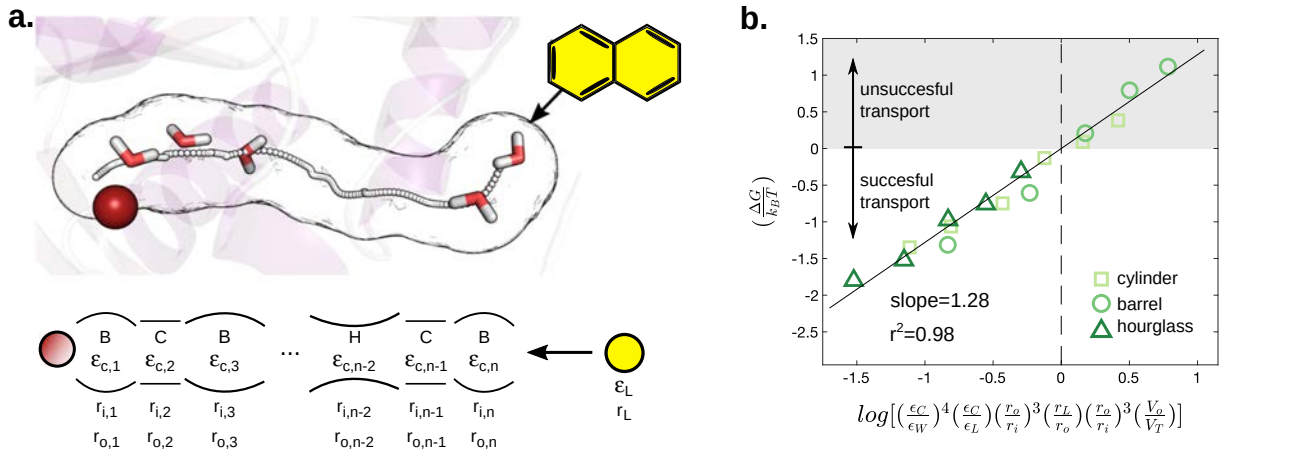

Fig. 1. Discretization of an enzyme nanochannel for the construction and mapping of the building block model. a. (top) Cartoon representation of naphthalene 1,2-dioxygenase (NDO) showing the surface of the channel wall (black), centerline of the channel (white dots), the mononuclear iron at the active site (red sphere), water molecules solvating the inside of the channel, and naphthalene (yellow) as the representative ligand. (bottom) Cartoon representing discretization of the NDO channel into the mapped building blocks. Each building block shows a schematic of the possible coarse-grained geometries, based on r_i and r_o, and the nonbonded interaction strength (ε) describing the level of wall hydrophobicity (see Fig. S1 for details). The ligand of interest (yellow circle) is represented by a spherical molecule of uniform hydrophobicity. b. Non-linear regression analysis relating dimensionless free energy to characteristic hydrophobicity, geometry, and excluded volume of the building block/ligand combination. The gray region shows the building block geometry and nonbonded interactions for which ligands did not successfully get transported through the building block; thus resulting in an unsuccessful transport.

binding to the buried active site of NDO is necessary for catalysis. Since ligands *must* overcome the geometric and/or energetic barriers imposed by the ~17 Å long channel to reach the active site [8], any positive catalytic activity can be used as a proxy for successful ligand transport through the channel. We performed two 40 ns molecular dynamics (MD) simulations for the unbound structure of NDO to study the effect of water on the geometry and hydrophobicity of the channel (to model wet vs. dry conditions), and thus on ligand transport. All simulation frames (time steps) were analyzed with Mole 2.0 [12], and an ensemble of one hundred frames with an open channel configuration were selected. The output of Mole 2.0 describes the enzyme channel in terms of geometric (radius) and physicochemical (hydrophobicity index – H_I) parameters. Details of the output parameters can be found in the methods section. For each frame, we obtained the channel radius and hydrophobicity index profile at 0.2 Å increments. The discrete values for both properties were interpolated and smoothed by a piecewise polynomial fitting spline (Fig. S3).

To understand the role water hydrogen bond (HB) networking plays inside the channel, we calculated the average ensemble radius for the wet and dry cases and found that the root-mean-squared-variation between the two states was <1 Å (see methods section for details of the analysis). This shows that, unlike some cytochrome P450s [13], the water molecules inside the NDO channel do not induce any major conformational changes on the channel architecture that may facilitate the transport of ligands. Instead, the displacement of these water molecules by the ligand from inside the channel to the bulk solvent phase may be the principal source of free energy controlling the transport into the active site. We also calculated the ensemble average H_I index for the wet channel and observed that the solvent exposed entrance region was less hydrophobic than the barrel-shaped mid-section of the channel (Fig. S3 bottom) as expected.

To test the validity of our ligand transport trajectory model, the splines for every frame were used to map each channel onto the sequence of building blocks. In the validation study a set of 45 ligands that have been previously tested experimentally [14], and computationally [8] in an all-atom model for catalytic activity in NDO were used. Fig. 2 shows a sample trajectory analysis for six different compounds in a single channel snapshot. For a molecule to have a non-zero probability of being catalyzed by NDO, it must first reach the active site [8,15]. In our model, this corresponds to the ligand having a value of $\Delta G^* < 0$ at all locations along the channel. Fig. 2 shows that naphthalene, isochroman, and diphenyl sulfide successfully completed transport through the full path to the active site of the enzyme, staying below the cutoff value at all times, and therefore were categorized as "possible substrates" of NDO, matching the experimentally observed results [14]. The detailed mechanism of inhibition by 1H-indole-3-acetate is still unknown, however, it has been proposed that the carboxylic group coordinates to the iron center [16]; meaning that this molecule must also reach the active site in order to inhibit NDO. The results presented in Fig. 2 show that the inhibitor was also able to reach the active site of the enzyme, matching the experimental observation. Finally, in the case of fluoranthene and 9,10-dihydro-9,10-methanoanthracene it was observed that only some portions of the trajectory were below the cutoff value, therefore they were correctly categorized as "unlikely substrates."

The trajectory analysis was repeated for all 100 frames to determine if the test ligands reached the active site. In Table S2 we report the probability of entrance into the active site for every ligand tested. We identified a very distinct pattern where the experimentally verified substrates of NDO were successfully transported to the active site through the building block model in >90% of the analyzed frames. On the other hand, the experimentally verified poor substrates only completed successful trajectories in <10% of the analyzed frames. Overall, our prediction accuracy was 90%, the positive prediction value was 90%, and the negative prediction value was 92% (Fig. S4 shows the prediction success rate at different discretization intervals). These values are slightly lower than the ones we observed in our previous computational studies [8]. However, the major benefit of this new method is the very fast channel transport prediction time of ~1 ms/ligand. This is a reduction in computation time of up to 6 orders of magnitude compared to our previously developed all-atom method [8]; Table S3 shows the prediction times for different currently available methods.

The improvement in computation time for our new method is the result of not having to perform the calculation-intensive all-atom MD simulations for every ligand. Instead, the new method utilizes the pre-calculated non-dimensional free energies (ΔG^*) needed to determine if the transport of a ligand along the set of interlocking building blocks will be favorable or unfavorable. These results show that the new method can be applied as a rapid pre-screening tool before any detailed, yet computationally expensive, all-atom methods is utilized. The

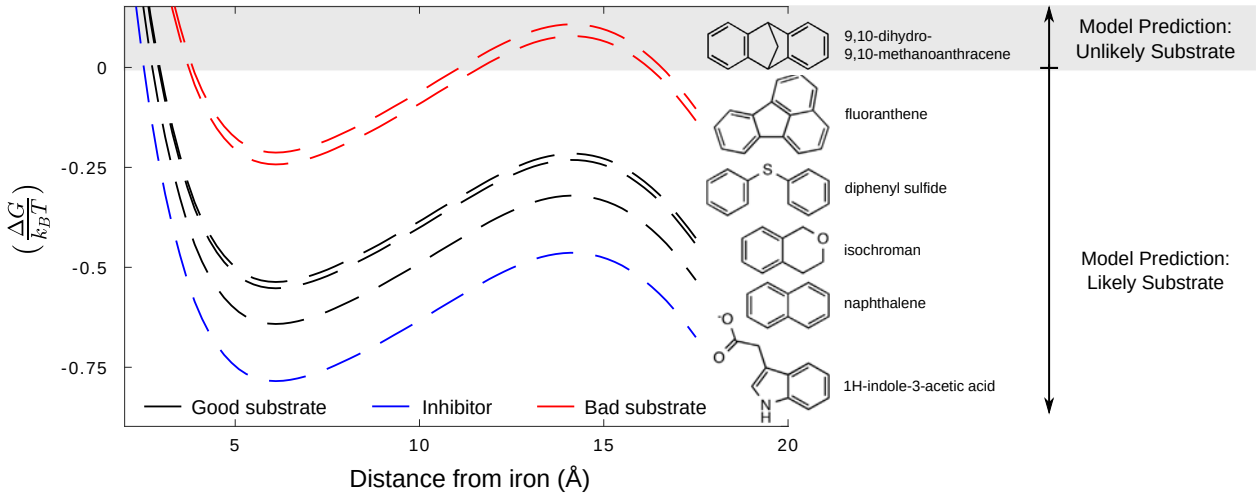

Fig. 2. Application of the building block trajectory analysis for six different compounds as shown for a single channel snapshot. Three known substrates (black lines) and one inhibitor (blue line) have successfully been transported along the building block model and reached the active site region (d < 6 Å), thus predicted to be likely substrates. Two poor substrates (red lines) have unfavorable trajectories at the bottleneck region (d = 12-17 Å), thus predicted to be unlikely substrates.

simplicity of the mapping procedure also allows the extension of this method to other fields, such as to analyze drug metabolism mediated by the network of hydrophobic channels found in some cytochrome P450s enzymes [17]. Overall, the approach presented here appears to be robust, transferable to other hydrophobic enzyme channels, and capable to elucidate the major geometric and energetic barriers that ligands experience as they move towards buried active sites. We expect that this method will be a valuable tool for the rational prediction of novel substrates for the production of biofuels, food and agricultural additives, and pharmaceuticals.

2. Methods

2.1. Molecular Dynamics and Tunnel Identification

We ran two 40 ns molecular dynamic simulation for the unbound structure of naphthalene 1,2-dioxygenase (PDB: 1O7G); details of the setup can be found elsewhere [8]. A total of one hundred snapshots were chosen at random from the last 20 ns of the production simulation, these snapshots represent the collective motion of the enzyme domains as well as the breathing motion of the channel leading into the active site. The channels connecting the solvent exposed area and the mononuclear iron were identified using the program Mole 2.0 [12]. Each of the channels identified by Mole 2.0 can be described by a set of coordinates detailing the centerline of the path connecting the iron (starting point, $d = 0$) and the bulk solvent. Hereafter, we call each individual point along the path of the channel a *step* and identify it by the subscript *k*. Similarly, at each step *k* the program returns the hydrophobicity index (H_I) describing the physicochemical environment of the channel. A similar simulation was performed for the unbound structure of NDO but having placed water molecules inside the channel using the "Solvate Pocket" feature of Desmond [18].

2.2. Conversion of Hydrophobicity Index

Mole 2.0 calculates the bulk hydrophobicity character of an enzyme channel using the scale proposed by Cid et al. [19] The value of H_I ranges between −1.2 and 1.85, positive values indicate non-polar hydrophobic regions, and negative values indicate polar and hydrophilic regions. In the development of our building block model we used the Lennard-Jones potential well ratio, normalized with respect to that of a SPC/E water molecule, to numerically describe the hydrophobicity of the building block wall surface. To determine a conversion factor between H_I and the potential well ratio used in our building block model, we correlated the hydrophobicity index with a mass-averaged LJPW for each amino acid. This mass-averaged LJPW value for each type of amino acid was calculated using Eq. (2):

$$\varepsilon_{AA} = \frac{\sum_i m_i \varepsilon_i}{\sum_i m_i} \tag{2}$$

where *i* refers to every atom in the amino acid and, *m* is the mass of the atom, and ε is the LJPW for that atom as parameterized by the OPLS force field.

2.3. Discretization of the Enzyme Channel

To construct our building block model we partitioned the enzyme channels into a set of *n* characteristic geometries each with a specific hydrophobicity (Fig. 1a). We first fitted a piecewise polynomial spline on the radius output from Mole 2.0 for all points *k* along the centerline of the channel. The spline was used to calculate the radius – r_o – of the channel at a distance *d* from the starting point as well as the radius – r_i – at a distance $d + 0.2$ Å from the starting point. The ratio r_i/r_o determines the type of geometry (i.e. barrel, cylinder or hourglass BB) and follows the same nomenclature used in the development of our building block model (Fig. S1) [9]. Similarly, we fitted a cubic spline on the H_I output from Mole 2.0 for all points *k* along the centerline of the channel. The calculated H_I spline was converted to the potential well ratio scale. The potential well ratio of the building block was calculated as the geometric average of the potential well ratio between points r_o and r_i. This procedure was repeated for every snapshot, resulting a total of 100 building block models.

2.4. Determination of Active Site Entry

Each ligand (small organic molecule) was modeled as a sphere of uniform hydrophobicity. The radius of the ligand sphere – r_L – was calculated from the solvent accessible surface area (A_S) for each molecule, i.e. $r_L = \sqrt{A_S/4\pi}$. And the potential well ratio for the ligand sphere was calculated by normalizing the mass averaged Lennard-Jones potential well of the molecule (using Eq. (2)) to the value of a SPC/E water molecule, i.e. $\varepsilon_L/\varepsilon_w$. The calculated characteristic parameters were used to determine ΔG^* for each building block/ligand combination, as described in

the "Nondimensinalization of building block model" section. These values were used to construct a ligand trajectory plot along each of the 100 enzyme channels. If at any point along the ligand trajectory plot $\Delta G^* > 0$, it was determined that the ligand did not enter the active site. Hence the condition of ligand entry was defined as $\Delta G^* < 0$ for all building blocks.

2.5. Nondimensionalization of Building Block Model

We constructed a *master plot* that can determine if a ligand is likely to successfully enter, and move along, and finally exit each *n* building block, based on previous thermodynamic and kinetic results. We used four characteristic nondimensional parameters to describe any building block/ligand combination: i) $\varepsilon_C/\varepsilon_W$, ii) $\varepsilon_C/\varepsilon_L$, iii) r_O/r_i, and iv) V_O/V_T. The last term is the excluded volume inside each building block at a given potential well ratio, this value can be directly obtained from *Table S1*. These four parameters were non-linearly fitted against the nondimensionalized Gibbs' free energy (ΔG^*), where $\Delta G^* = \Delta G/k_BT$.

2.6. Nondimensional Gibb's Free Energy

The relationship between Gibb's free energy (ΔG) and the equilibrium constant (K_{eq}),

$$\Delta G = -k_B T \ \ln\left(K_{eq}\right) \tag{3}$$

was nondimensionalized resulting in:

$$\Delta G/k_B T = \Delta G^* = \ \ln\left(1/K_{eq}\right), \tag{4}$$

where k_B is the Boltzmann constant, and T is the temperature of the system.

The transport of ligands along the building blocks depend on the geometric and physicochemical parameters, as well as the excluded volume fraction. Therefore, the equilibrium constant of our transport process must also be a function of on these parameters. We expressed the equilibrium constant K_{eq} as a function of 25 nondimensional geometric and physicochemical parameters in our system in the following form:

$$\frac{1}{K_{eq}} = f\left(\alpha^a \beta^b \gamma^c \ldots \omega^z\right), \tag{5}$$

inserting Eq. (5) into Eq. (4) results in:

$$\Delta G^* = \ \ln\left(\alpha^a \beta^b \gamma^c \ldots \omega^z\right). \tag{6}$$

we then estimated the value of the exponents a, b, c…z using Matlab's iterative least-squared non-linear regression solver (*nlinfit* function). Majority of the exponents were zero, resulting in the equation:

$$\Delta G^* = 0.55 \ \ln\left[\left(\frac{\varepsilon_C}{\varepsilon_W}\right)^4 \left(\frac{\varepsilon_C}{\varepsilon_L}\right)^{1.5} \left(\frac{r_o}{r_i}\right)^3 \left(\frac{V_O}{V_T}\right)\right], \tag{7}$$

or:

$$\Delta G^* = 1.28 \ \log\left[\left(\frac{\varepsilon_C}{\varepsilon_W}\right)^4 \left(\frac{\varepsilon_C}{\varepsilon_L}\right)^{1.5} \left(\frac{r_o}{r_i}\right)^3 \left(\frac{V_O}{V_T}\right)\right] \tag{8}$$

in $\log_{10}$ basis.

The first two terms inside the logarithm, of Eqs. (7) and (8), represent the energetic contributions to the equilibrium constant, whereas the last two terms represent the entropic contributions due to the transport of the ligand through the building blocks. In a favorable process where a ligand gets transported through the building block $\Delta G^* < 0$. Therefore, $\Delta G^* = 0$ is defined as the cut-off between the favorable and unfavorable transport processes, resulting in Eq. (1).

Acknowledgements

This research was funded through a University of Minnesota IonE Discovery grant, a fellowship (to DE) from the University of Minnesota Informatics Institute, and the MnDRIVE Transdisciplinary Initiative of the University of Minnesota. The authors acknowledge the Minnesota Supercomputing Institute (http://www.msi.umn.edu) at the University of Minnesota for providing resources that contributed to the research results presented in this paper.

Declarations of Interests

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2019.06.001.

References

[1] Kaushik S, et al. Impact of the access tunnel engineering on catalysis is strictly ligand-specific. FEBS J 2018;285:1456–76. https://doi.org/10.1111/febs.14418.
[2] Pravda L, et al. Anatomy of enzyme channels. BMC Bioinformatics 2014;15:379. https://doi.org/10.1186/s12859-014-0379-x.
[3] Wheeldon I, et al. Substrate channelling as an approach to cascade reactions. Nat Chem 2016;8:299.
[4] Brezovsky J, et al. Software tools for identification, visualization and analysis of protein tunnels and channels. Biotechnol Adv 2013;31:38–49. https://doi.org/10.1016/j.biotechadv.2012.02.002.
[5] Macarron R, et al. Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov 2011;10:188–95. https://doi.org/10.1038/nrd3368.
[6] Paloncyova M, Navratilova V, Berka K, Laio A, Otyepka M. Role of Enzyme flexibility in ligand access and egress to active site: Bias-Exchange metadynamics study of 1,3,7-Trimethyluric Acid in Cytochrome P450 3A4. J Chem Theory Comput 2016;12:2101–9. https://doi.org/10.1021/acs.jctc.6b00075.
[7] Rydzewski J, Nowak W. Ligand diffusion in proteins via enhanced sampling in molecular dynamics. Phys Life Rev 2017;22-23:58–74. https://doi.org/10.1016/j.plrev.2017.03.003.
[8] Escalante DE, Aukema KG, Wackett LP, Aksan A. Simulation of the bottleneck controlling access into a Rieske active site: predicting substrates of naphthalene 1,2-Dioxygenase. J Chem Inf Model 2017;57:550–61. https://doi.org/10.1021/acs.jcim.6b00469.
[9] Escalante DE, Aksan A. Role of water hydrogen bonding on transport of small molecules inside hydrophobic channels. J Phys Chem B 2019 [submitted].
[10] Bassan A, Blomberg MRA, Siegbahn PEM. A theoretical study of the cis-dihydroxylation mechanism in naphthalene 1,2-dioxygenase. J Biol Inorg Chem 2004;9:439–52. https://doi.org/10.1007/s00775-004-0537-0.
[11] Wolfe MD, Parales JV, Gibson DT, Lipscomb JD. Single turnover chemistry and regulation of O2 activation by the oxygenase component of naphthalene 1,2-dioxygenase. J Biol Chem 2001;276:1945–53. https://doi.org/10.1074/jbc.M007795200.
[12] Sehnal D, et al. Mole 2.0: advanced approach for analysis of biomacromolecular channels. J Chem 2013;5:39. https://doi.org/10.1186/1758-2946-5-39.
[13] Hendrychova T, Berka K, Navratilova V, Anzenbacher P, Otyepka M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. Curr Drug Metab 2012;13:177–89.
[14] Aukema KG, Kasinkas L, Aksan A, Wackett LP. Use of silica encapsulated *Pseudomonas sp.* strain NCIB 9816-4 in biodegradation of novel hydrocarbon ring structures found in hydraulic fracturing waters. Appl Environ Microbiol 2014;80:4968–76. https://doi.org/10.1128/AEM.01100-14.
[15] Prokop Z, et al. Engineering of protein tunnels: Keyhole-lock-key model for catalysis by the enzymes with buried active sites. Wiley-VCH; 2012; 421–64.
[16] Karlsson A, et al. NO binding to naphthalene dioxygenase. J Biol Inorg Chem 2005;10:483–9. https://doi.org/10.1007/s00775-005-0657-1.
[17] Lonsdale R, Rouse SL, Sansom MSP, Mulholland AJA. Multiscale approach to modelling drug metabolism by membrane-bound cytochrome P450 enzymes. PLoS Comput Biol 2014;10. https://doi.org/10.1371/journal.pcbi.1003714 doi:ARTN e1003714.
[18] Wang LL, Berne BJ, Friesner RA. Ligand binding to protein-binding pockets with wet and dry regions. P Natl Acad Sci USA 2011;108:1326–30. https://doi.org/10.1073/pnas.1207504109.
[19] Cid H, Bunster M, Canales M, Gazitua F. Hydrophobicity and structural classes in proteins. Protein Eng 1992;5:373–5. https://doi.org/10.1093/protein/5.5.373.