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 of geometries and hydrophobicity of the system, as well as exclusion volume values for each ligand/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 non-dimensional Gibbs free energy of transport (ΔG* = ΔG/kBT) and the dimensionless parameters that characterizes the contributions of hydrophobicity and geometry of the system, as well as exclusion volume effects inside the building blocks.

\[ ΔG^* = \log \left( \frac{\varepsilon_c}{\varepsilon_W} \right)^{\frac{4}{3}} \left( \frac{r_0}{r_1} \right)^{\frac{15}{3}} \left( \frac{V_0}{V_T} \right)^{\frac{3}{3}} = 0 \]  

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...
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 ($\Delta G^*$) 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
simplex 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 additivites, 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 \((H_k)\) 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” 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 \( H_k \) 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 \( H_k \) 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):

\[
\varepsilon_{iA} = \sum_i \frac{m_i \varepsilon_i}{m} 
\]

where \( i \) refers to every atom in the amino acid and, \( m \) is the mass of the atom, and \( \varepsilon \) 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 – \( r_0 \) – of the channel at a distance \( d \) from the starting point as well as the radius – \( r_1 \) – at a distance \( d + 0.2 \) Å from the starting point. The ratio \( r_0/r_1 \) 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 \( H_k \) output from Mole 2.0 for all points \( k \) along the centerline of the channel. The calculated \( H_k \) 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 \( r_0 \) and \( r_1 \). 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 – \( r_0 \) – was calculated from the solvent accessible surface area \( (A_{SA}) \) for each molecule, i.e. \( r_0 = \sqrt{A_{SA}/4\pi} \). 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. \( \varepsilon_{SPC}/\varepsilon_{water} \). The calculated characteristic parameters were used to determine \( \Delta G^0 \) for each building block/ligand combination, as described in

![Image](https://example.com/image.png)
the “Nondimensionalization 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) $e_v/e_w$, ii) $e_v/e_t$, iii) $r_0/r_s$, and iv) $V_0/V_T$. 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 Gibb’s free energy ($\Delta G^*$), where $\Delta G^* = \Delta G/k_BT$.

2.6. Nondimensional Gibb’s Free Energy

The relationship between Gibb’s free energy ($\Delta G$) and the equilibrium constant ($K_{eq}$),

$$\Delta G = -k_BT \ln (K_{eq}) \quad (3)$$

was nondimensionalized resulting in:

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

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:

$$K_{eq}^{-1} = f\left(\alpha^a \beta^b \gamma^c \cdots \omega^q\right). \quad (5)$$

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

$$\Delta G^* = \ln \left(\alpha^a \beta^b \gamma^c \cdots \omega^q\right). \quad (6)$$

we then estimated the value of the exponents $a, b, c, \ldots$ using Matlab’s iterative least-squared non-linear regression solver ($nlinf$t function). Majority of the exponents were zero, resulting in the equation:

$$\Delta G^* = 0.55 \ln \left[\left(\frac{e_v}{e_w}\right)^4 \left(\frac{e_v}{e_t}\right)^{1.5} \left(\frac{r_0}{r_s}\right)^3 \frac{V_0}{V_T}\right]. \quad (7)$$

or:

$$\Delta G^* = 1.28 \log \left[\left(\frac{e_v}{e_w}\right)^4 \left(\frac{e_v}{e_t}\right)^{1.5} \left(\frac{r_0}{r_s}\right)^3 \frac{V_0}{V_T}\right]. \quad (8)$$

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).