Three-dimensional quantitative structure–activity relationship and docking studies in a series of anthocyanin derivatives as cytochrome P450 3A4 inhibitors

Sergey Shityakov\(^1\)
István Puskás\(^2\)
Norbert Roewer\(^1\)
Carola Förster\(^1\)
Jens Broscheit\(^1\)

\(^1\)Department of Anesthesia and Critical Care, University of Würzburg, Würzburg, Germany; \(^2\)CycloLab Cyclodextrin Research and Development Laboratory Ltd, Budapest, Hungary

Abstract: The cytochrome P450 (CYP)3A4 enzyme affects the metabolism of most drug-like substances, and its inhibition may influence drug safety. Modulation of CYP3A4 by flavonoids, such as anthocyanins, has been shown to inhibit the mutagenic activity of mammalian cells. Considering the previous investigations addressing CYP3A4 inhibition by these substances, we studied the three-dimensional quantitative structure–activity relationship (3D-QSAR) in a series of anthocyanin derivatives as CYP3A4 inhibitors. For the training dataset (n=12), comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) yielded crossvalidated and non-crossvalidated models with a $q^2$ of 0.795 (0.687) and $r^2$ of 0.962 (0.948), respectively. The models were also validated by an external test set of four compounds with $r^2$ of 0.821 (CoMFA) and $r^2$ of 0.812 (CoMSIA). The binding affinity modes associated with experimentally derived IC\(\text{_{50}}\) (half maximal inhibitory concentration) values were confirmed by molecular docking into the CYP3A4 active site with $r^2$ of 0.66. The results obtained from this study are useful for a better understanding of the effects of anthocyanin derivatives on inhibition of carcinogen activation and cellular DNA damage.

Keywords: three-dimensional quantitative structure–activity relationship, cytochrome P450 3A4, comparative molecular field analysis, comparative molecular similarity index analysis, anthocyanin derivatives, molecular docking, carcinogen activation

Introduction

Cytochrome P450 (CYP) proteins play important roles in metabolism in many species. These enzymes are involved in the detoxification of not only endogenous molecules, but also various exogenous substances, such as biogenic amines,\(^1\) cytostatic agents,\(^2\) and steroids.\(^3\) A substantial fraction of the CYP family is CYP3A, which comprises up to 30\% of the total liver CYP enzyme pool in humans.\(^4\) Moreover, its 3A4 isoenzyme is believed to be responsible for approximately 60\% of metabolic drug changes.\(^5\)

Notwithstanding the membrane-associated state of mammalian CYPs and difficulties in their crystallization, significant progress has been made recently in this direction by obtaining precise and detailed information about CYP structures.\(^6\)-\(^8\) In addition, a number of three-dimensional quantitative structure–activity relationship (3D-QSAR) studies, including comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA), have been performed on the different CYP inhibitors to predict, correlate, and interpret experimentally determined inhibition values and assess substrate selectivity.\(^9\)-\(^11\) These methods are based on the assumption that the changes in
inhibition activity of chemical substances are related to changes in structural properties represented by molecular fields.\textsuperscript{12} Several flavonoids have been shown to modulate the CYP system and alter the pharmacokinetics and pharmacodynamics of many commonly prescribed drug substances.\textsuperscript{13,14} These water-soluble glycosidic derivatives of flavylum salts are most prevalent in various fruit and vegetables, such as cabbage, grapes, and berries.\textsuperscript{15} It was previously reported by Dreisitl et al that CYP3A4 inhibition by a series of anthocyanin derivatives, including their aglycones and proanthocyanidins, is achieved in a concentration-dependent manner.\textsuperscript{14} However, it was suggested by the same authors that their inhibitory potencies were adversely correlated with the number of sugar moieties responsible for a decrease in the effects of anthocyanidins on CYP3A4.\textsuperscript{14}

The antimutagenic activity of anthocyanidins was earlier established by Gasiorowski et al, whose results revealed that these compounds, isolated from the fruit of\textit{Aronia melanocarpa}, sufficiently inhibit the mutagenic activity of benzo(a) pyrene and 2-aminofluorene in the Ames test.\textsuperscript{16} Hence, CYP inhibition by anthocyanin compounds might be considered as antimutagenic, given that these enzymes contribute to carcinogenesis in several organs\textsuperscript{17-20} and increase DNA toxicity and damage.\textsuperscript{21}

Therefore, in order to evaluate a series of anthocyanin derivatives as CYP3A4 inhibitors, two computational approaches were used in the present study, ie, 3D-QSAR and molecular docking techniques, to analyze and match experimentally derived CYP3A4 half maximal inhibitory concentration (IC\textsubscript{50}) values to their predicted ones together with the binding affinity modes. To our knowledge, this work presents the first 3D-QSAR and docking methods that might be of interest to determine the most appropriate CoMFA and CoMSIA prediction models for selective CYP inhibition and to block the mutagenic and carcinogenic activity mediated by CYP activation mechanisms.

**Computational methods**

The chemical formulae of 16 anthocyanin derivatives were obtained from research papers published elsewhere.\textsuperscript{23,24} The two-dimensional structures of the compounds were drawn using the MarvinSketch version 5.2 program (ChemAxon, Budapest, Hungary). After the two-dimensional sketches were converted into three-dimensional images, explicit hydrogen molecules were added and recorded in PDB (Protein Data Bank) format; 3D-QSAR modeling was done using Sybyl-X 1.1 software (Tripos International, St Louis, MO, USA). Rigid body alignment of anthocyanin derivatives was performed using a maximum common substructure by the PyMol software (Schrödinger LLC, Rockville, MD, USA)\textsuperscript{22} to build the CoMFA and CoMSIA models. CoMSIA similarity indices ($A_{qK}$) for a molecule $j$ with atoms $i$ at a grid point $q$ were calculated as follows:

$$A_{qK}^{i,j} = \sum_{i=1}^{n} \omega_{\text{probe},i} K_{ar,i,q} e^{-\alpha q^2}$$

where $K$ represents the following physicochemical properties: steric and electrostatic descriptors; $\omega_{\text{probe},i}$ indicates probe atom; $\alpha$ is an attenuation factor; $K_{ar,i,q}$ is the actual value of the physicochemical property $K$ of atom $i$; and $q^2$ is the mutual distance between the probe atom at grid point $q$, and atom $i$ of the test molecule.

The compounds were divided into a training dataset (compounds 1–12, Figure 1) and test dataset (compounds 13–16, Figure 2) using a random selection method that is part of the Strike 1.9 module integrated in the Maestro 9.1 molecular modeling suite (Schrödinger LLC, Rockville, MD, USA). A 75% compound separation irrespective of inhibition activity or structural composition was achieved for the whole dataset. Gasteiger–Hückel charges were added before the molecular field calculations and model elaboration. Energy minimization was performed using the Tripos force-field\textsuperscript{23,24} with a distance-dependent conjugate gradient algorithm. The minimum gradient of 0.05 kcal*mol$^{-1}$ was set as a convergence criterion.\textsuperscript{24} Default values were used with a 2.0 Å grid spacing using the common sp$^3$ carbon atom hybridization with a +1 point charge as a probe to investigate the steric and electrostatic interactions at the lattice grid points. Meanwhile, a default value of 0.3 was used as the attenuation factor, and a distance-dependent Gaussian-type functional form was used between the grid point $q$ and each atom $i$ in the molecular structure. This setting can avoid singularities at the atomic positions and the dramatic changes of potential energy caused by the grid in surface proximity.\textsuperscript{25} The column filtering value was set to 2.0 kcal*mol$^{-1}$. Statistical analysis was performed in the Sybyl-X program (Tripos International, St Louis, MO, USA) using the partial least square method. Crossvalidated ($q^2$) and non-crossvalidated ($r^2$) values were determined for the models using linear regression analysis with variances represented as the standard error of estimation.\textsuperscript{26} For the number of components, a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecutive default (six components) was provided in the a consecut
The electron diffraction of human microsomal CYP3A4 as a holostructure (PDB ID: 1TQN) at 2.05 Å resolution with the heme cofactor was retrieved from the Protein Data Bank as a PDB file. Structural refinement and molecular docking studies were performed on the CYP3A4 crystal structure using DockingServer (Virtua Drug Ltd, Budapest, Hungary), which allows for the carrying out of highly efficient and robust docking calculations in one comprehensive web service. The ligand preparation included Gasteiger partial charges assignment and rotatable bonds definition before all nonpolar hydrogen atoms were merged. All necessary solvation parameters were added with the aid of

Figure 1 Chemical structure of anthocyanin derivatives used as a training set for the three-dimensional quantitative structure–activity relationship study.
Rigid-flexible molecular docking was applied to the center of the heme-associated binding cavity using the following Cartesian coordinates: \( x = -15.67 \, \text{Å}, \, y = -22.49 \, \text{Å}, \) and \( z = -11.2 \, \text{Å} \). The AutoDock output results represented the docking scores as Gibbs free energy of binding (\( \Delta G \)) values, further converted to the predicted inhibition constants \( (K_{i_{(\text{pred})}}) \). Affinity grid maps of \( 20\times20\times20 \, \text{Å} \) grid points, and 0.375 Å spacing were generated using the AutoGrid program (Scripps Research Institute, San Diego, CA, USA). AutoDock parameter set-dependent and distance-dependent dielectric functions were used in the calculation of the van der Waals and electrostatic terms, respectively. Docking simulations were performed using the Lamarckian genetic algorithm and the Solis and Wets local search method. Initial position, orientation, and torsions of the ligand molecules were set randomly. All rotatable torsions were released during docking. Each docking experiment was derived from ten different runs that were set to terminate after a maximum of 250,000 energy evaluations while the population size was set to 150. During the search, a translational step of 0.2 Å, and quaternion and torsion steps of 5 were also applied.

The \( IC_{50} \) values were obtained from the study reported by Dreiseitel et al, as determined when investigating the inhibition of CYP3A4 activity in a concentration-dependent manner using a chemiluminescent assay. The biological data was further converted to \( pK_{i_{(\text{exp})}} \) (negative decimal logarithm of experimental inhibition constant) parameters using the \( pIC_{50} \) value calculator according to the Michaelis–Menten enzyme–substrate interaction equation for noncompetitive species:

\[
K_i = IC_{50}, \quad \text{when} \quad S = K_m \text{ or } S \gg K_m \text{ or } S \ll K_m
\]

where \( S \) is the substrate concentration and \( K_m \) is the Michaelis–Menten constant.

The statistical parameters of the standard CoMFA model were constructed using rigid body alignment with all charges assigned to the molecules. To emphasize the differences between the minimal-energy conformations of anthocyanin derivatives, we aligned these compounds with the pelargonidin structure used as a common scaffold, and measured the root-mean-square deviation values. The root-mean-square deviation parameters significantly correlated in an inverse manner with the experimentally determined inhibitory constants \( (pK_{i_{(\text{exp})}}) \); structural deviation decreased

**Results and discussion**

The statistical parameters of the standard CoMFA model were constructed using rigid body alignment with all charges assigned to the molecules. To emphasize the differences between the minimal-energy conformations of anthocyanin derivatives, we aligned these compounds with the pelargonidin structure used as a common scaffold, and measured the root-mean-square deviation values. The root-mean-square deviation parameters significantly correlated in an inverse manner with the experimentally determined inhibitory constants \( (pK_{i_{(\text{exp})}}) \); structural deviation decreased
with an increase in compound activity and vice versa (Figure 3A–C). The most deviated substances (root-mean-square deviation >0.1 Å), such as cyanidin-3,5-diglucoside, cyanidin-3-rutinoside, cyanidin-3-galactoside, malvidin-3,5-diglucoside, and peonidin-3-glucoside, had additional sugar moieties in their structure that mainly contributed to the conformational mismatch. The sugar residues changed their spatial orientation as a result of significant conformational rearrangements caused by rotations around the ω1 (C2-C3-O-sugar) and ω2 (C6-C5-O-sugar) dihedral angles.

The $q^2$, $r^2$, $F$, and standard error of estimation values were computed as defined in the Sybyl-X suite. The training set was preliminarily checked for outliers. Examination of the residuals from the crossvalidated predictions indicated that there were no outliers in the CoMFA model. According to the results from the CoMFA analysis shown in Tables 1 and 2, the crossvalidated partial least square ($q^2$) was found to be 0.795 with five components, indicating that the CoMFA model is statistically significant and highly predictive. The non-crossvalidated regression value showed a conventional $r^2$ of 0.962, an $F$ of 250.8, and a standard error of estimation of 0.083 ($P<$0.0001), suggesting that the model could correctly predict compound activity in 96% of cases (Figure 4A). The contributions from the steric and electrostatic fields were

![Figure 3](image_url)

**Figure 3** (A) Rigid body alignment of the anthocyanin compounds used in this study. Hydrogen atoms are omitted for clarity. A bar chart for the compound numbers (B) and schematic data smoothing (C) were utilized to compare the values and produce a fitting curve for the RMSD and pK$_{i\ (exp)}$ comparison.

**Abbreviations:** RMSD, root-mean-square deviation; pK$_{i\ (exp)}$, negative decimal logarithm of experimental inhibition constant.

<table>
<thead>
<tr>
<th></th>
<th>CoMFA</th>
<th>CoMSIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-validated</td>
<td>$q^2$</td>
<td>0.795</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.687</td>
</tr>
<tr>
<td>Number of components</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Column filtering (kcal/mol)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Non-crossvalidated</td>
<td>$r^2$</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>SEE</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>$F$</td>
<td>250.8</td>
</tr>
<tr>
<td></td>
<td>Steric</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>Electrostatic</td>
<td>0.324</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.557</td>
</tr>
</tbody>
</table>

**Table 1** Results from the three-dimensional quantitative structure–activity relationship analysis for the training dataset.
Table 2 Predicted affinities for CoMFA and CoMSIA models of anthocyanin derivatives

<table>
<thead>
<tr>
<th>Compound</th>
<th>IC(_{50}) (μM)</th>
<th>pK(_i) (exp)</th>
<th>pK(_i) (CoMFA)</th>
<th>Residual</th>
<th>pK(_i) (CoMSIA)</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyanidin</td>
<td>20.8</td>
<td>4.68</td>
<td>4.722</td>
<td>0.042</td>
<td>4.694</td>
<td>0.014</td>
</tr>
<tr>
<td>Cyanidin-3,5-diglucoside</td>
<td>167.5</td>
<td>3.76</td>
<td>3.781</td>
<td>0.021</td>
<td>3.804</td>
<td>0.058</td>
</tr>
<tr>
<td>Cyanidin-3-galactoside</td>
<td>80.2</td>
<td>4.095</td>
<td>4.197</td>
<td>0.102</td>
<td>4.193</td>
<td>0.098</td>
</tr>
<tr>
<td>Cyanidin-3-glucoside</td>
<td>105.3</td>
<td>3.98</td>
<td>4.075</td>
<td>0.095</td>
<td>4.058</td>
<td>0.078</td>
</tr>
<tr>
<td>Cyanidin-3-rutinoside</td>
<td>96.1</td>
<td>4.02</td>
<td>4.022</td>
<td>0.002</td>
<td>3.997</td>
<td>-0.02</td>
</tr>
<tr>
<td>Delphinidin-3-glucoside</td>
<td>74.1</td>
<td>4.13</td>
<td>4.174</td>
<td>0.044</td>
<td>4.083</td>
<td>-0.047</td>
</tr>
<tr>
<td>Malvidin-3,5-diglucoside</td>
<td>225.5</td>
<td>3.65</td>
<td>3.561</td>
<td>-0.089</td>
<td>3.572</td>
<td>-0.078</td>
</tr>
<tr>
<td>Malvidin-3-galactoside</td>
<td>84.3</td>
<td>4.07</td>
<td>4.09</td>
<td>0.002</td>
<td>4.206</td>
<td>0.136</td>
</tr>
<tr>
<td>Malvidin-3-glucoside</td>
<td>80</td>
<td>4.096</td>
<td>3.99</td>
<td>-0.106</td>
<td>4.005</td>
<td>-0.091</td>
</tr>
<tr>
<td>Pelargonidin</td>
<td>12.2</td>
<td>4.91</td>
<td>4.777</td>
<td>-0.133</td>
<td>4.75</td>
<td>-0.16</td>
</tr>
<tr>
<td>Peonidin</td>
<td>16</td>
<td>4.795</td>
<td>4.721</td>
<td>-0.074</td>
<td>4.729</td>
<td>-0.066</td>
</tr>
<tr>
<td>Petunidin</td>
<td>23.4</td>
<td>4.63</td>
<td>4.706</td>
<td>0.076</td>
<td>4.746</td>
<td>0.116</td>
</tr>
<tr>
<td>Delphinidin</td>
<td>46.5</td>
<td>4.333</td>
<td>4.697</td>
<td>0.364</td>
<td>4.658</td>
<td>0.325</td>
</tr>
<tr>
<td>Malvidin</td>
<td>14</td>
<td>4.85</td>
<td>4.709</td>
<td>-0.141</td>
<td>4.72</td>
<td>-0.13</td>
</tr>
<tr>
<td>Pelargonidin-3,5-diglucoside</td>
<td>249</td>
<td>3.603</td>
<td>3.611</td>
<td>0.008</td>
<td>3.737</td>
<td>0.134</td>
</tr>
<tr>
<td>Peonidin-3-glucoside</td>
<td>79.2</td>
<td>4.101</td>
<td>3.993</td>
<td>-0.108</td>
<td>3.954</td>
<td>-0.147</td>
</tr>
</tbody>
</table>

Abbreviations: CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity index analysis; IC\(_{50}\), half maximal inhibitory concentration; pK\(_i\) (exp), negative decimal logarithm of experimental inhibition constant; pK\(_i\) (CoMFA), negative decimal logarithm of CoMFA-predicted inhibition constant; pK\(_i\) (CoMSIA), negative decimal logarithm of CoMSIA-predicted inhibition constant.

Figure 4 (A) Predicted and experimental pK\(_i\) values from the comparative molecular field analysis model correlated with each other by linear regression analysis. The comparative molecular field analysis maps are shown in respect to pelargonidin as a reference molecule. (B) Steric maps are colored in green for steric bulk areas with an increased activity. (C) Electrostatic maps are colored in blue for positively charged areas with an increased activity and red ones for negatively charged areas. Hydrogen atoms are omitted for clarity.

Abbreviations: pK\(_i\) (exp), negative decimal logarithm of experimental inhibition constant; pK\(_i\) (CoMFA), negative decimal logarithm of CoMFA-predicted inhibition constant.

detected to be 67% and 33%, respectively. This observation was unexpected given that the anthocyanin derivatives used to build up the CoMFA model were all positively charged. Therefore, their positive charges contribute mainly to electrostatic fields but not to the steric hindrance, which might be due to the volume restrictions associated with the narrow binding cavities.

In the next step, the CoMFA model was validated by the external test set of four compounds, ie, delphinidin, malvidin, pelargonidin-3,5-diglucoside, and peonidin-3-glucoside.
It can be seen from Tables 2 and 3 that the CoMFA model could correctly predict the compound affinity with an $r^2$ of 0.821, an $F$ of 9.117, and a standard error of estimation of 0.281 ($P<0.0939$), suggesting that the CoMFA model could accurately predict compound activity in 82% of cases. The pelargonidin molecule as the strongest CYP3A4 inhibitor (IC$_{50}$ 12.2 µM) is shown in Figure 4B and C with the CoMFA steric and electrostatic maps. The green maps represent the areas where increasing bulk would likely lead to an increase in activity, while the yellow steric regions (if any) indicate areas with decreased compound activity. We observed that there is a zone of optimal steric favor that surrounds the hydroxyl groups of 4-hydroxyphenyl and chromenyl-3,5,7-triol moieties. The blue and red electrostatic maps indicate the positively and negatively charged areas with an increase and decrease in activity, respectively. Interestingly, the large red area was located at the 4-hydroxyphenyl part of the molecule, indicating only minor contribution to the inhibition effect due to its conformational flexibility. This flexibility might be explained by the phenomenon of aglycone internal rotation around the $\theta$ angle (O1-C2-C1’-C2’), which defines the coplanarity of the A, B, and C rings. In general, the maps represent a molecular state that is close to optimal for binding to the CYP3A4 protein when it was detected as IC$_{50}$ in the previous experiments.$^{14}$

The results of the CoMSIA analysis are shown in Tables 1 and 2. Similarly to the CoMFA model, we used the steric and electrostatic fields in the CoMSIA analysis. Partial least square crossvalidation showed a regression value ($q^2$) of 0.687, which indicates a good predictive model. The optimal number of components (five) was found from previous crossvalidation analysis to build up the nonvalidated linear regression model with a high Pearson’s correlation coefficient ($r^2$) of 0.948, an $F$ of 184.2, and a standard error of estimation of 0.095 ($P<0.0001$, Figure 5A). The relative contributions from the steric and electrostatic fields

![Figure 5](https://www.dovepress.com/)

**Figure 5 (A)** Predicted and experimental pK$_{I}$ values from the comparative molecular similarity index analysis model correlated with each other by linear regression analysis. The comparative molecular similarity index analysis maps are shown in respect to pelargonidin as a reference molecule. (B) Steric maps are colored in green for steric bulk areas with an increased activity and yellow ones, where steric bulk will likely lead to a decrease in activity. (C) Electrostatic maps are colored in blue for positively charged areas with increased activity and red ones for negatively charged areas. Hydrogen atoms are omitted for clarity.

**Abbreviations:** pK$_{I}$ (exp), negative decimal logarithm of experimental inhibition constant; pK$_{I}$ (CoMSIA), negative decimal logarithm of CoMSIA-predicted inhibition constant.
were detected to be 44% and 56%, respectively. In this case, all molecular charges contributing to the electrostatic fields were dominating over the steric effect. The latter parameter was decreased as a result of the relatively large size of the CYP3A4 binding cavity that might release the steric hindrance in various sizes and shapes of ligand structures.

It can also be seen from Tables 2 and 3 that the CoMSIA model was validated by the external test set of the same four compounds with an $r^2$ of 0.812, an $F$ of 8.636, and a standard deviation error of estimation of 0.263 ($P<=0.0989$), suggesting that the CoMSIA model could accurately predict compound activity in 81% of cases.

As compared with the CoMSIA model, the CoMFA model appeared to be more predictive, partly due to the small deviation molecular alignment and also the fact that CoMSIA relies on explicit hydrophobic and hydrogen bond donor or acceptor field descriptors.

The CoMSIA model is shown in Figure 5B and C with the pelargonidin structure used as a reference molecule. The green maps indicate areas where addition of bulky steric chemical groups would presumably impose an increase in compound activity, while the yellow ones would reverse this effect, as already observed in the previous CoMFA model. Additionally, there was an increase in positively charged areas with higher activity colored in blue in the vicinity of the hydroxyl group in the 4-hydroxyphenyl moiety and the hydroxyl group in the 3'-position of the C-ring of the flavonoid backbone.

To gain some insight into the binding characteristics of anthocyanin derivatives as CYP3A4 inhibitors, we docked them into the heme cofactor-associated binding cavity of CYP3A4. During the docking process, the protein was considered to be rigid while the ligands were flexible. The structure of CYP3A4 turned out to be much more open in the proximity of the heme ion, resulting in more uniform distribution of the cavity volume and easier ligand binding.27

Eventually, only 40% of the total binding cavity area was occupied by all docked compounds, reflecting the relatively large size of the active cavity, which is a well established characteristic of CYP3A4.6,54 Thus, this enhances simultaneous molecular occupancy, which occurs near the catalytic center of the enzyme that contributes to heterotrophic cooperativity by facilitating alternative binding modes for CYP3A4 inhibitors, as shown experimentally in many studies.35–37 However, anthocyanin derivatives occupy different binding regions within the same binding cavity (Figure 6A), with different amino acid residues involved because of their dihedral rotation and conformational mismatch.

In order to find common similarities in the CYP3A4 ligand interaction modes for the strongest inhibitors, we selected pelargonidin as the best experimental CYP3A4 inhibitor with a $pK_{i\text{exp}}$ of 4.91 and cyanidin as the best predicted CYP3A4 binder with a $pK_{i\text{pred}}$ of 5.21. Taken together, we found that the Phe447 and Phe435 aromatic rings of

![Figure 6](image-url)

**Figure 6** (A) Rigid-flexible docking of anthocyanin derivatives into the binding site of the cytochrome P450 3A4 molecule. The molecular surface is divided by the frontal plane to visualize a binding channel of the protein structure. Red and blue colors are depicted for negative and positive electrostatic potentials; while zero potential is in white. Ligands are colored according to their atom types. Hydrogen atoms are omitted for clarity. (B) Relationship between predicted binding constants calculated from the docking studies with cytochrome P450 3A4 and experimental ones determined by a sensitive chemiluminescence assay for 16 anthocyanin derivatives.

**Abbreviations:** $pK_{i\text{exp}}$, negative decimal logarithm of experimental inhibition constant; $pK_{i\text{pred}}$, negative decimal logarithm of predicted inhibition constant.

### Table 3 Results from the three-dimensional quantitative structure–activity relationship analysis for the test dataset

<table>
<thead>
<tr>
<th></th>
<th>CoMFA</th>
<th>CoMSIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^2$</td>
<td>0.821</td>
<td>0.812</td>
</tr>
<tr>
<td>SEE</td>
<td>0.281</td>
<td>0.263</td>
</tr>
<tr>
<td>$F$</td>
<td>9.117</td>
<td>8.636</td>
</tr>
</tbody>
</table>

**Abbreviations:** SEE, standard error of estimation; CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity index analysis.
CYP3A4 interact with these two molecules by establishing π-π stacking with the A ring for pelargonidin and the B ring for cyanidin (Figure 7A and B). Similarly, previous theoretical and experimental studies have elucidated the effect of π-π stacking between phenylalanine groups and aromatic rings of various CYP3A4 substrates and inhibitors.\(^\text{38,39}\) In addition, the cyanidin hydroxyl functional group at position 4 of the B ring (3,4-dihydroxyphenyl moiety) forms an H-bond with the backbone acceptor oxygen atom of Pro434, thereby contributing to an increase in binding affinity.

In the molecular docking procedure, a cysteine amino acid (residue number 442) presents at the binding pocket, which is a strong nucleophile and participates easily in covalent interactions with other compounds. Furthermore, there are some reports in the literature of cysteine residues in the CXXCH motif of CYP acting to form covalent thioether bonds.\(^\text{40,41}\) Therefore, there might be some opportunity for a covalent interaction to occur between the Cys442 residue and the ligands.

The next step was to explore whether or not the experimental values (K\(_i\)\(_{\text{exp}}\) or pK\(_i\)\(_{\text{exp}}\)) can correlate with the ΔG-derived inhibition constants (K\(_i\)\(_{\text{pred}}\) or pK\(_i\)\(_{\text{pred}}\)) (Table 4). The Michaelis–Menten enzyme–substrate interaction equation was applied to the inhibitory activities represented as IC\(_{50}\) values. Consequently, a model with \(r^2\) of 0.66 was obtained for 16 compounds (Figure 6B), which indicates that the correlation was statistically significant, for the theoretical molecular docking study, thus demonstrating the applicability of this method. This model eventually confirmed the previous observations found in our CoMFA/CoMSIA models and previously published experimental results\(^\text{14,42}\) that anthocyanin derivatives, generally occurring as 3-O-glycosides, have bioactivity attributable to the aglycone but not to the sugar moiety or other binding species.

### Conclusion

In this paper, we describe the development of 3D-QSAR models with CoMFA and CoMSIA methods and molecular docking procedures to explain the inhibition of anthocyanin derivatives with CYP3A4.
derivatives on CYP3A4. Initial attempts at using a ligand-based method for construction of the CoMFA and CoMSIA models were based on the 16 reported anthocyanin derivatives through additions of hydroxyl groups and glycoside residues in different positions and this resulted in models with over 80% predictability. Switching to the protein-based docking method, the predicted binding energy from AutoDock provided a correlation with the experimental IC50 values of at least 66%. The π-π stackings with Phe447 and Phe435 were indicated as key interactions for the anthocyanin derivatives. The hydrogen bond interaction between Pro434 and the hydroxyl group on the cyanidin B ring was also considered to contribute to the increased binding affinity.

Overall, in the present investigation, the number of sugar residues predicted and confirmed a “structural prerequisite” for a decline in the anthocyanin inhibition effect on CYP3A4, which underscores the urgent need for more detailed analysis of the structure–activity relationship.

Finally, we emphasize the importance of structure–activity relationship models in the development of CYP inhibitors. These models could be a valuable tool in the design of drug-like molecules, which are targeted for use as inhibitors of cytochromes, which are the enzymes responsible for carcinogen activation and cellular DNA damage.

Acknowledgments

Special thanks are extended to Dr Nigel Aminake from Aachen University and Anna Poon from the City College of New York for their assistance in the writing of this paper. The authors are grateful to the Bundesministerium für Bildung und Forschung for its support of this work by providing grants (BMBF01, EO1004) to CF.

Disclosure

The authors report no conflicts of interest in this work.

References


