

## Original Article

# Novel chemical scaffolds of the tumor marker AKR1B10 inhibitors discovered by 3D QSAR pharmacophore modeling

Raj KUMAR<sup>1</sup>, Minky SON<sup>1</sup>, Rohit BAVI<sup>1</sup>, Yuno LEE<sup>1</sup>, Chanin PARK<sup>1</sup>, Venkatesh ARULALAPPERUMAL<sup>1</sup>, Guang Ping CAO<sup>1</sup>, Hyong-ha KIM<sup>2</sup>, Jung-keun SUH<sup>3</sup>, Yong-seong KIM<sup>4</sup>, Yong Jung KWON<sup>5</sup>, Keun Woo LEE<sup>1,\*</sup>

<sup>1</sup>Division of Applied Life Science (BK21 Plus Program), Systems and Synthetic Agrobiotech Center (SSAC), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Research Institute of Natural Science (RINS), Gyeongsang National University (GNU), 501 Jinju-daero, Jinju 660–701, Republic of Korea; <sup>2</sup>Division of Quality of Life, Korea Research Institute of Standards and Science, Daejeon 305–340, Republic of Korea; <sup>3</sup>Bio Computing Major, Korean German Institute of Technology, Seoul 157–033, Republic of Korea; <sup>4</sup>Department of Science Education, Kyungnam University, Masan 631–701, Republic of Korea; <sup>5</sup>Department of Chemical Engineering, Kangwon National University, Chunchon 200–701, Republic of Korea

**Aim:** Recent evidence suggests that aldo-keto reductase family 1 B10 (AKR1B10) may be a potential diagnostic or prognostic marker of human tumors, and that AKR1B10 inhibitors offer a promising choice for treatment of many types of human cancers. The aim of this study was to identify novel chemical scaffolds of AKR1B10 inhibitors using *in silico* approaches.

**Methods:** The 3D QSAR pharmacophore models were generated using HypoGen. A validated pharmacophore model was selected for virtual screening of 4 chemical databases. The best mapped compounds were assessed for their drug-like properties. The binding orientations of the resulting compounds were predicted by molecular docking. Density functional theory calculations were carried out using B3LYP. The stability of the protein-ligand complexes and the final binding modes of the hit compounds were analyzed using 10 ns molecular dynamics (MD) simulations.

**Results:** The best pharmacophore model (Hypo 1) showed the highest correlation coefficient (0.979), lowest total cost (102.89) and least RMSD value (0.59). Hypo 1 consisted of one hydrogen-bond acceptor, one hydrogen-bond donor, one ring aromatic and one hydrophobic feature. This model was validated by Fischer's randomization and 40 test set compounds. Virtual screening of chemical databases and the docking studies resulted in 30 representative compounds. Frontier orbital analysis confirmed that only 3 compounds had sufficiently low energy band gaps. MD simulations revealed the binding modes of the 3 hit compounds: all of them showed a large number of hydrogen bonds and hydrophobic interactions with the active site and specificity pocket residues of AKR1B10.

**Conclusion:** Three compounds with new structural scaffolds have been identified, which have stronger binding affinities for AKR1B10 than known inhibitors.

**Keywords:** AKR1B10 inhibitor; tumor marker; molecular docking; molecular dynamics; quantitative structure-activity relationship; virtual screening; density functional theory

Acta Pharmacologica Sinica (2015) 36: 998–1012; doi: 10.1038/aps.2015.17; published online 8 June 2015

## Introduction

Aldo-keto reductase family 1 B10 (AKR1B10) is an NADPH-dependent oxidoreductase belonging to the human superfamily of aldo-keto reductase (AKR) genes. AKR1B10 is a monomeric protein that is mainly expressed in the human colon, the small and large intestines, the adrenal glands and at low levels

in the liver<sup>[1,2]</sup>. The expression of AKR1B10 in the human gastrointestinal epithelium is related to steroid metabolism and the prevention of carbonyl lesions<sup>[3]</sup>. Earlier, AKR1B10 was isolated from hepatocellular carcinoma (HCC) and was more recently found to be overexpressed in many systemic malignancies, including squamous cell pulmonary carcinomas, pancreatic carcinomas, bladder carcinogenesis and colorectal carcinomas<sup>[4]</sup>. The expression of AKR1B10 has been correlated with lung cancer and smoking habits. Cigarette smoke and tobacco consumption induces AKR1B10 expression, which

\* To whom correspondence should be addressed.

E-mail kwlee@gnu.ac.kr

Received 2014-12-01 Accepted 2015-02-27

has been proposed as a new diagnostic marker for smoking-related non-small cell lung carcinoma (NSCLC)<sup>[5]</sup>. Specific substrates of AKR1B10 include farnesal, geranylgeranyl, retinal and carbonyls. Reduced products of farnesal and geranylgeranyl are intermediates of cholesterol synthesis involved in protein prenylation, which appears to be a crucial event in carcinogenesis that regulates the activation of KRAS-mutations and the MAP kinase cascade, eventually leading to pancreatic cancer<sup>[2, 6]</sup>.

Retinoic acid is an important signaling molecule which directs cell differentiation. AKR1B10 can reduce retinal to retinol (thus limiting retinoic acid formation), therefore stimulating carcinogenesis by controlling the retinoid signaling pathway<sup>[7]</sup>. Under physiological conditions, highly reactive aldehyde and ketone groups lead to cell death by carbonyl-induced apoptosis. AKR1B10 plays a role in the detoxification of cytotoxic carbonyl compounds and promotes cell survival by controlling carbonyl metabolism<sup>[8]</sup>. Furthermore, AKR1B10 can reduce the carbonyl groups in some anticancer drugs to their corresponding alcohols, thereby decreasing the effectiveness of these drugs<sup>[9]</sup>. Therefore, AKR1B10 plays a crucial role in carcinogenesis by promoting cell survival and metastasis, and its importance as a potential cancer drug target has been well documented. The crystal structure of AKR1B10 represents the typical ( $\alpha/\beta$ )<sub>8</sub> barrel topology characteristic of the AKR superfamily proteins<sup>[1, 10]</sup>. The active site is located at the C-terminal end of the barrel and a binding pocket is formed between external loops B and C. The NADP<sup>+</sup> is adjoined near the active site in a stretched conformation. The residues Tyr49, His111, and Trp112 as well as the nicotinamide moiety of NADP<sup>+</sup> form the anion-binding pocket or inhibitor binding site. The residues belonging to loop A (Lys125 to Ala131) contribute to the upper lid of the active site pocket. The residues Trp21, Val48, Trp80, Trp112, Phe116, Phe123, Trp220, Cys299, Val301, and Gln303 as well as the nicotinamide moiety of the cofactor form a strong hydrophobic pocket that confers substrate specificity<sup>[1]</sup>.

Recently, many gene silencing studies of AKR1B10 gene have demonstrated the growth inhibition of cancer cells, emphasizing the use of AKR1B10 as a potential diagnostic or prognostic tumor marker<sup>[2, 6, 11-13]</sup>. AKR1B10 inhibitors offer a promising choice for treatment of many types of human cancers. Several synthetic and natural compounds have been reported to inhibit AKR1B10<sup>[1, 14-16]</sup>. AKR1B10 inhibitors may carry out the inhibition of the structurally similar human aldose reductase AKR1B1 in the AKR superfamily<sup>[1]</sup>. Therefore, it is important to design novel chemical scaffolds that can act as selective inhibitors of AKR1B10. 2-phenyliminochromene derivatives are the most potent known inhibitors of AKR1B10. Among them, 7-hydroxy-2-(4-methoxyphenylimino)-2H-chromene-3-carboxylic acid benzylamide is a selective inhibitor of AKR1B10 (IC<sub>50</sub>=4.7 nmol/L)<sup>[17]</sup> that has been used as reference compound in this study. Due to a low number of AKR1B10 inhibitors, there is a need to identify new scaffolds that are more selective and possess higher inhibitory activities. Therefore, we used computer-

aided drug design approaches to identify novel, potent inhibitors that can selectively inhibit AKR1B10. A 3D QSAR pharmacophore model was built from the chemical features present in already known inhibitors. The best model, Hypo 1, was validated and used for database screening. Binding conformations of the selected hit compounds were predicted using molecular docking studies. The potential compounds were filtered by checking their drug-like properties. The molecular orbital energies were calculated using density functional theory (DFT) and compared to assess the high reactivities of the compounds. Finally, the appropriate binding modes of the final hit compounds were revealed by molecular dynamics (MD) simulations studies.

## Materials and methods

### Selection of compounds and dataset preparation

The choice of a training set of compounds is critical for the generation of a pharmacophore model, which subsequently determines the quality of the generated pharmacophores. A total of 64 AKR1B10 inhibitors with a range of different inhibitory activities were obtained from reported publications<sup>[1, 13-15]</sup>. The compounds were divided into the training set and the test set. The training set was used to build the pharmacophore model and the test set was exploited for pharmacophore model validation. Twenty-four training set compounds were selected by satisfying the selection rule in *HypoGen*<sup>[18]</sup>. The training set compounds had diverse chemical structures and a range of activities (IC<sub>50</sub> values) that spanned four orders of magnitude. The remaining 40 compounds were used as the test set. The inhibitory activities (IC<sub>50</sub>) of the training set compounds ranged from 4.7 nmol/L to 90000 nmol/L. The compounds of the training set were categorized as active (IC<sub>50</sub><100 nmol/L, +++), moderately active (100 nmol/L≤IC<sub>50</sub><10000 nmol/L, ++) and inactive (IC<sub>50</sub>≥10000 nmol/L, +) based on their IC<sub>50</sub> values. Similarly, the compounds of the test set were also classified based on their activities. The 2D chemical structures of all compounds were drawn using ChemSketch<sup>[19]</sup> and were subsequently exported to Discovery Studio v3.5 (DS) for the generation of their corresponding 3D structures.

### Pharmacophore model generation

The 3D QSAR Pharmacophore Generation protocol available in DS was used to generate predictive pharmacophores. This program uses the Catalyst *HypoGen* algorithm<sup>[20]</sup> to generate hypotheses from common chemical features in a training set of compounds with known activity values (IC<sub>50</sub>). Low energy conformations of the compounds were generated using the *BEST* algorithm. The energy threshold value was set to 20 kcal/mol<sup>[21]</sup>. The uncertainty value, which represents the ratio of the uncertainty range of the actual activity against the measured biological activity for each compound, was kept at 3. The other parameters were kept at their default values. The *Feature Mapping* protocol in DS was used to carefully investigate the important chemical features of the training set compounds. The mapped chemical features such as hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), ring

aromatic (RA) and hydrophobic regions (HYP) were used to generate the hypotheses. The minimum and maximum number of all the features in the hypotheses tested were set to 0 and 5, respectively. Ten quantitative hypotheses were generated with their corresponding statistical parameters, which included the cost values (null and fixed costs), correlation ( $R^2$ ), root mean square deviation (RMSD), and fit values. The cost values were analyzed as per Debnath's method<sup>[22]</sup>.

### Hypothesis validation

The general requirements of a hypothesis included its statistical significance, its accurate prediction of the activity of the compounds, and its ability to retrieve active compounds from the chemical databases. The best pharmacophore model was validated to assess its ability to discriminate active compounds from inactive compounds. The selected pharmacophore model was evaluated for its cost values, assessed using Fischer's method, and validated using the test set of compounds. The statistical significance of a hypothesis was calculated as

$$S=1-\left(\frac{1+X}{Y}\right)\times 100$$

where 'X' represents the total number of hypotheses with a total cost value lower than the most significant hypothesis and 'Y' is the number of initial *HypoGen* runs plus random runs<sup>[21]</sup>. Fischer's randomization method checks the correlation between the chemical structure and the biological activity of a compound. This method overrules the probability of a chance correlation for pharmacophore model development and ensures that the model was not generated randomly. The confidence level was set to 95% in the 3D QSAR pharmacophore generation process. As a result, 19 random spreadsheets were automatically generated by DS. The test set was used to determine whether the generated pharmacophore hypothesis could predict and classify the compounds according to their ranges of experimental activities. Low energy conformations were generated using the same protocols used for the training set compounds. The *Ligand Pharmacophore Mapping* module of DS was used with the *BEST* algorithm and the *Flexible* fitting option.

### Virtual screening and drug-likeness prediction

Database screening was conducted to identify novel compounds as potential AKR1B10 inhibitors. Pharmacophore-based database searching is a type of ligand-based virtual screening that can be used to find novel and potential leads for further drug development. A potent pharmacophore model possesses the chemical functionalities responsible for the bioactivities of potential drugs, thus suggesting its use in performing a database search. The validated quantitative pharmacophore model was used as a 3D query to screen four different chemical databases: NCI, Asinex, Chembridge, and Maybridge. A molecule contained within a database should map all features of the pharmacophore model to be retrieved as a hit. The *Ligand Pharmacophore Mapping* protocol of DS was used for database screenings with *Fast* and *Flexible* options.

The compounds that fit all the features of the best pharmacophore model were retrieved as hits. To ensure drug-like physicochemical properties, the hit compounds were filtered by applying Lipinski's rule of five<sup>[23]</sup>. This rule suggests that a drug is well-absorbed when the compound has less than 10 hydrogen bond acceptor groups, less than 5 hydrogen bond donor groups, a molecular weight of less than 500 Da, a Log *P* value of less than 5, and less than 10 rotatable bonds. The absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of each compound were calculated using the *ADMET Descriptors* protocol in DS. The compounds that fulfilled the drug-likeness properties were chosen for molecular docking studies.

### Molecular docking

The molecular docking of screened ligands and the target protein has emerged as a very effective tool in the modern drug discovery process<sup>[24]</sup>. This method can be used to monitor the interactions and behavior of small molecules in the binding site of target proteins. Here, the aim of the docking study was to predict the binding modes of hit compounds and estimate their binding affinities. The training set compounds and 125 hit compounds were chosen for molecular docking. The docking studies were carried out using GOLD v5.2.2 software (Genetic Optimization for Ligand Docking)<sup>[25, 26]</sup>. The crystal structure of AKR1B10 (PDB code: 4I5X)<sup>[10]</sup> was downloaded from the protein data bank (www.rcsb.org). All water molecules were removed and hydrogen atoms were added to calculate the bond orders for the protein and ligand. The orientations of all histidine tautomers were transformed into the ND1H protonation states, as in the crystal structure<sup>[10]</sup>. The binding site of the protein was defined for all the atoms within 10 Å of the co-crystallized ligand in the crystal structure. Goldscore was used as the default scoring function to predict the binding affinity of the ligand to the target protein, while rescoring was conducted using Chemscore. Ten docking poses were generated for each ligand and the best poses were selected based on high Goldscores. Furthermore, the docked poses were analyzed based on the molecular interactions and formation of hydrogen bonds between the ligand and the active site residues.

### Density functional theory

DFT is a promising approach that has been efficiently utilized in the quantum mechanical simulation of periodic systems<sup>[27]</sup>. DFT provides an accurate explanation of the electronic and structural properties of small molecules by computing the electronic structure of matter. The orbital energies associated with frontier orbitals provide a useful means of characterizing the electron donor and acceptor properties of the molecules, which confer their overall reactivity. Most chemical reactions of a molecule are governed by these orbitals. The energy of the highest occupied molecular orbitals (HOMO) is directly related to the ionization potential (electron donor), whereas the lowest unoccupied molecular orbitals (LUMO) are related to the electron affinity (electron acceptor)<sup>[19, 28, 29]</sup>. The HOMO

and LUMO are indicators of the possible electrophilic and nucleophilic attack sites in the molecules, respectively. The selected docked poses of the hit compounds were used as the inputs for the DFT calculations. The B3LYP exchange-correlation potential functional of Dmol3 in DS was employed for calculating the orbital energies. The calculation was performed to compare the electronic properties of the hits to those of the compounds in the training set.

### Molecular dynamics simulations

The final hit compounds obtained from the docking and DFT calculations as well as the most active compounds in the training set were subjected to 10 ns molecular dynamics (MD) simulations. The MD simulations were carried out with a CHARMM27 all-atom force field using GROMACS 4.5.7 package<sup>[30]</sup>. The topology files for the cofactors and ligands were generated using the online web service SwissParam (<http://www.swissparam.ch/>)<sup>[31]</sup>. An octahedral water box of 1 nm thickness was generated and solvated using a TIP3P water model. The system was neutralized by replacing water molecules with Na<sup>+</sup> counter-ions. The energy of the system was minimized through 10000 steps using the steepest descent algorithm to remove all possible bad contacts with a maximum force lower than 1000 kJ/mol. After the energy minimization, the NVT equilibration was conducted for 200 ps at 300 K. A V-rescale thermostat was used to maintain constant temperature. NVT was followed by 200 ps NPT equilibration applied at a pressure of 1 bar. These equilibrated structures were used to perform 10 ns MD simulations at 300 K (maintained by a V-rescale thermostat)<sup>[32]</sup> and 1 bar (maintained by a Parrinello-Rahman barostat)<sup>[33]</sup>. During equilibration, the protein backbone was restrained and the solvent molecules along with counter-ions were allowed to move. The LINCS algorithm<sup>[34]</sup> was employed to restrain the bonds of heavy atoms and their corresponding hydrogen atoms. The particle mesh Ewald (PME) method<sup>[35]</sup> was used to calculate the long-range electrostatic interactions and a cut-off value of 12 Å

was applied to measure the short-range interactions. The MD simulations were performed under periodic boundary conditions to avoid edge effects<sup>[36]</sup>. The simulations were conducted with a time step of 2 fs and the coordinate data were stored to the file every 1 ps. The results were analyzed using the GROMACS<sup>[30]</sup>, VMD<sup>[37]</sup>, and DS software. The representative structure, which is the closest conformation to the average structure, was selected from each simulation and used for the analysis.

### Binding free energy calculations of AKR1B10 and reference inhibitor/hit compounds

The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method was used to compute the binding free energies of the protein-ligand complexes, as described earlier<sup>[38-40]</sup>. For this purpose, 20 snapshots of each complex (AKR1B10 and reference inhibitor/hit compounds) were chosen evenly from 0 to 10 ns MD trajectories according to earlier reports<sup>[38, 41]</sup>. The obtained snapshots were used to calculate the energetic parameters using the MM/PBSA method<sup>[38, 41, 42]</sup>.

## Results

### Pharmacophore model generation

The training set compounds were used to generate pharmacophore models using the *HypoGen* algorithm (Figure 1). A total of ten hypotheses were generated using the statistical parameter values such as cost values (null and fixed costs), R<sup>2</sup>, RMSD, and fit values (Table 1). Statistically, the best hypothesis has the highest cost difference, lowest total cost values, high correlation coefficient, and smallest RMSD values<sup>[23]</sup>. The difference between the null cost and total cost of each hypothesis is designated as the cost difference. A hypothesis may have a high correlation coefficient if the cost difference is greater than 60. A low cost difference score (less than 40) of a hypothesis represents a correlation below 50%, which makes it difficult to find a significant predictive model.

Hypo 1 contains four chemical features, including HBA,

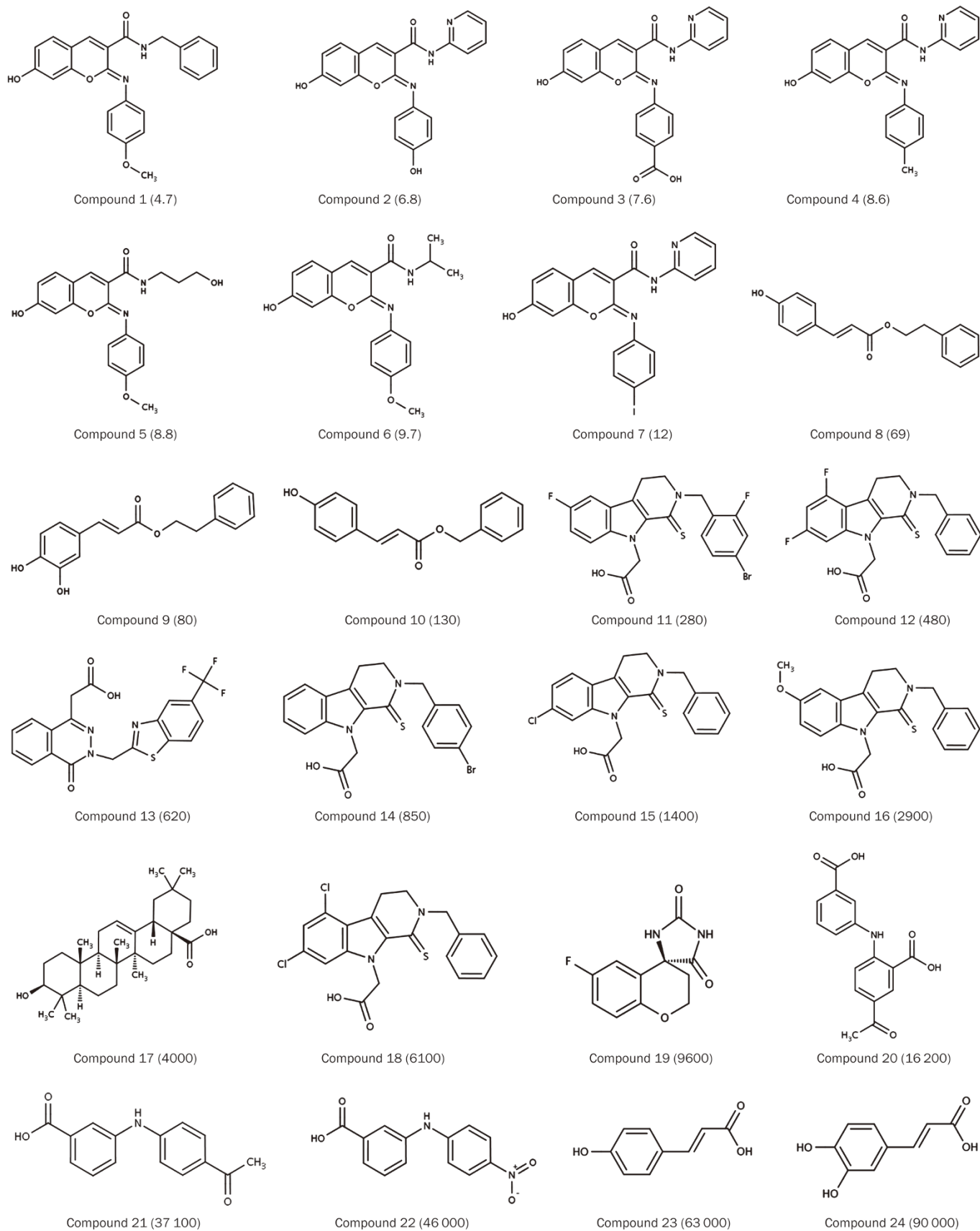
**Table 1.** Statistical results of ten pharmacophore hypotheses generated by *HypoGen*.

| Hypo No | Total cost | Cost difference <sup>a</sup> | RMSD <sup>b</sup> | Correlation (R <sup>2</sup> ) | Max fit | Features <sup>c</sup> |
|---------|------------|------------------------------|-------------------|-------------------------------|---------|-----------------------|
| Hypo 1  | 102.89     | 83.33                        | 0.59              | 0.979                         | 8.74    | HBA, HBD, RA, HYP     |
| Hypo 2  | 102.92     | 83.29                        | 0.61              | 0.978                         | 8.31    | HBA, HBD, RA, HYP     |
| Hypo 3  | 103.66     | 82.56                        | 0.66              | 0.975                         | 7.91    | HBA, HBD, RA, HYP     |
| Hypo 4  | 103.79     | 82.43                        | 0.66              | 0.974                         | 8.56    | HBA, HBD, RA, HYP     |
| Hypo 5  | 104.15     | 82.07                        | 0.69              | 0.972                         | 8.28    | HBA, HBD, RA, HYP     |
| Hypo 6  | 104.47     | 81.75                        | 0.71              | 0.971                         | 8.17    | HBA, HBD, RA, HYP     |
| Hypo 7  | 104.67     | 81.56                        | 0.72              | 0.969                         | 8.12    | HBA, HBD, RA, HYP     |
| Hypo 8  | 105.09     | 81.13                        | 0.72              | 0.970                         | 9.15    | HBA, HBD, RA, HYP     |
| Hypo 9  | 106.07     | 80.16                        | 0.79              | 0.963                         | 8.28    | HBA, HBD, RA, HYP     |
| Hypo 10 | 106.17     | 80.05                        | 0.79              | 0.963                         | 8.72    | HBA, HBD, RA, HYP     |

<sup>a</sup>Cost difference: difference between the null cost and the total cost. The null cost of ten scored hypotheses is 186.224, the fixed cost value is 98.402, and the configuration cost is 16.55. All costs are represented in bit units.

<sup>b</sup>RMSD: deviation of the log (estimated activities) from the log (experimental activities) normalized by the log (uncertainties).

<sup>c</sup>HBA: hydrogen bond acceptor; HBD: hydrogen bond donor; RA: ring aromatic; HYP: hydrophobic.



**Figure 1.** A representation of the 24 chemically diverse training set compounds used for pharmacophore generation. The experimental  $IC_{50}$  values (nmol/L) are shown in parentheses for each compound.

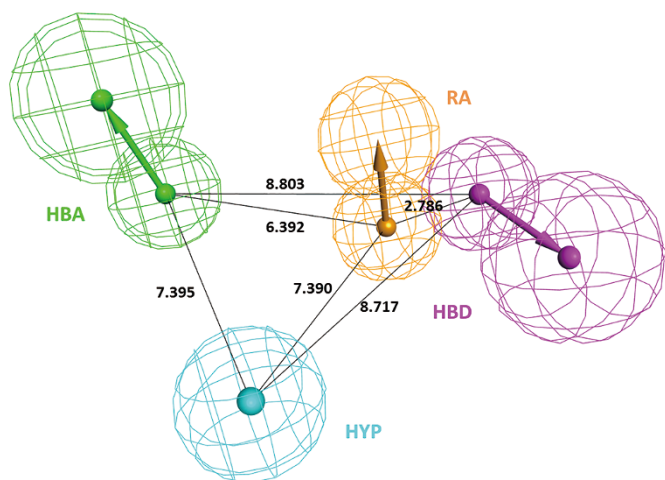
HBD, RA, and HYP, and was selected as the best pharmacophore model (Figure 2). The model showed the highest cost difference value (83.33), lowest RMSD (0.59), and maximum correlation (0.979). Additionally, this model showed the highest fit value of 8.74 compared to the other hypotheses (Table 1). The high correlation demonstrates the greater predictive ability of the model, whereas the lower RMSD value represents a low deviation of the predicted activity from the experimental activity. Therefore, Hypo 1 was used to estimate the inhibitory activities of 24 training set compounds and to interpret the predictive accuracy of this model. The most active ( $IC_{50}=4.7$  nmol/L) and least active ( $IC_{50}=90\,000$  nmol/L) compound in the training set both aligned with Hypo 1 (Figure 3). The most active compound mapped all the features in Hypo 1,

whereas the least active compound failed to map the HBA and HYP features. This demonstrates the difference in activities between the most active and the least active compounds. Hypo 1 was used to evaluate the activity of compounds with a high degree of accuracy in relation to their experimental  $IC_{50}$  values (Table 2). The inhibitory activity values of all active compounds were predicted in the same order of magnitude, although two moderately active compounds were estimated as active and inactive, respectively. The error value was defined as the ratio between the experimental and predicted activity values. The Hypo 1 model was forwarded to a validation analysis.

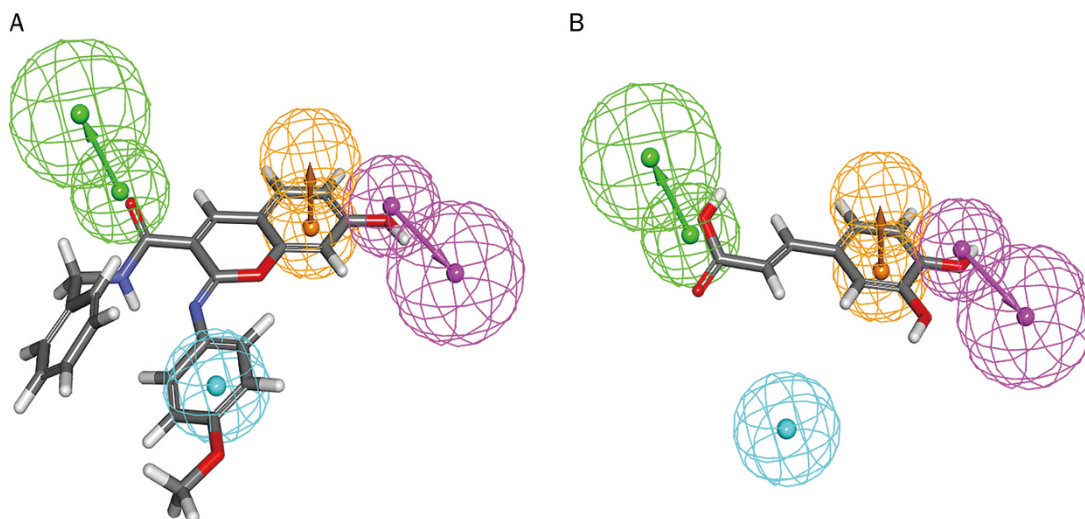
#### Hypothesis validation

The significance of each generated pharmacophore model was assessed by Fischer's randomization method and test set validation. In Fischer's randomization test, a confidence level of 95% was used. Consequently, a set of 19 random spreadsheets indicating the total cost values were calculated for ten generated hypotheses (Figure 4). The total cost values in the random spreadsheets were the least for Hypo 1 compared to the nine other hypotheses. This indicated that the Hypo 1 model was statistically significant and not generated by chance.

The test set validation was used to assess the ability of Hypo 1 to predict the activity of compounds and classify them into the correct activity range. The 40 known AKR1B10 inhibitors (Supplementary Table 1) in the test set were categorized as active, moderately active and inactive. Hypo 1 was able to classify all test set compounds according to their activity ranges except for one inactive compound, which was estimated as moderately active (Table 3). Hypo 1 showed strong correlation coefficients between the experimental and predicted activities in both the training and test set compounds (Figure 5). Thus, the test set validation results demonstrated that Hypo 1 can be used to discriminate the active compounds from the moderately active and inactive compounds.

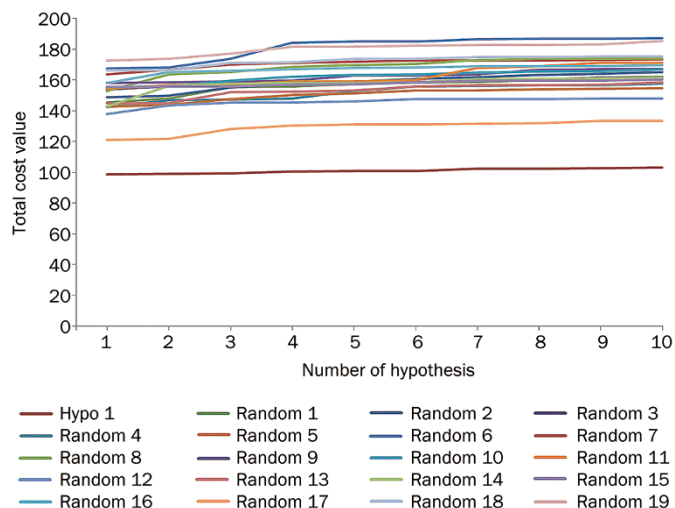


**Figure 2.** The best pharmacophore model, Hypo 1, with distance constraints. Hypo 1 contains one hydrogen bond acceptor (HBA: green), one hydrogen bond donor (HBD: magenta), one ring aromatic (RA: orange), and one hydrophobic region (HYP: cyan).

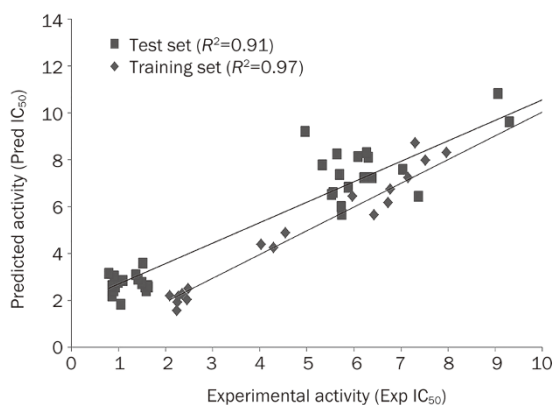


**Figure 3.** Alignment of Hypo 1 to training set compounds. (A) most active compound 1 ( $IC_{50}=4.7$  nmol/L) and (B) least active compound 24 ( $IC_{50}=90\,000$  nmol/L). The most active compound mapped to all four features in Hypo 1, whereas the least active compound missed the HBA and HYP features.



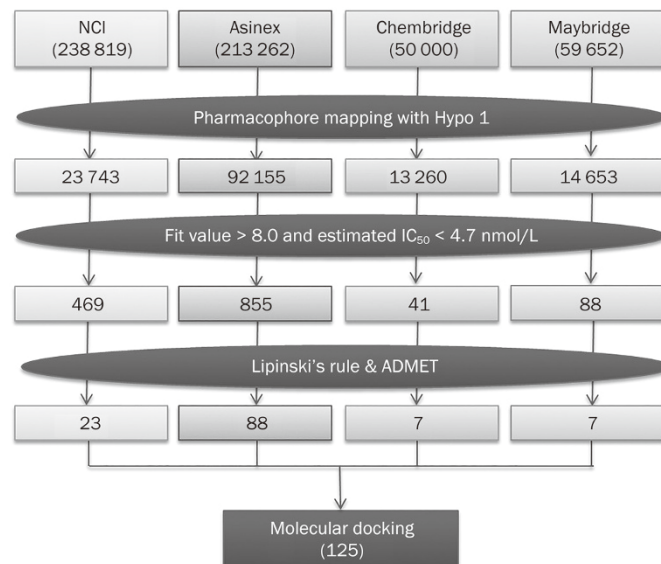


**Figure 4.** A graphical representation of the total cost values of Hypo 1 and each of ten hypotheses generated from 19 random spreadsheets during Fischer's randomization run. A confidence level of 95% was used.



**Figure 5.** Correlations between the experimental activities and the predicted activities using Hypo 1 with the test set and training set compounds.

docked pose showed an acceptable RMSD difference of 0.65 Å between the X-ray structure and the predicted structure (Supplementary Figure 1). Additionally, the reliability of the generated pharmacophore model was assessed by superimposing the docked pose of the reference compound to the conformation generated in HypoGen. An acceptable RMSD value (1.15 Å) was obtained after superimposition (Supplementary Figure 2). The most active compound from the training set was selected as the reference compound to compare the docking results of the other compounds. The Goldscore of the reference compound was 37.869. Therefore, a Goldscore of 38.0 was taken as cut-off for the further screening of compounds, and a total of 80 compounds were selected. The range of docking score for the compounds was 41.593 to 68.52. Chemscore was used as the rescoring function. While this program estimates the total free energy change that occurs upon ligand



**Figure 6.** The overall results of the virtual screening using Hypo 1.

binding, it is not superior to Goldscore in predicting affinities<sup>[43, 44]</sup>. Therefore, the compounds with the lowest Chemscores were chosen for further analysis. The Chemscore for the reference inhibitor and three hit compounds (hit 1, hit 2, and hit 3) were -34.41, -42.58, -42.51, and -38.07 kJ/mol, respectively. Furthermore, the compounds were checked manually for their interactions with the active site residues of AKR1B10. The reference inhibitor showed a similar binding mode to that reported by Endo *et al*<sup>[17]</sup>. A total of 30 compounds that have shown hydrogen bond interactions with active site residues such as Tyr49, His111, and Trp112 were chosen for calculating the molecular orbital energies by DFT.

#### Density functional theory

The energies of the HOMO and LUMO frontier orbitals are responsible for charge transfer in a chemical reaction. The HOMO and LUMO characterize the susceptibility of the molecule to attack by electrophiles and nucleophiles, respectively. A high HOMO energy corresponds to a high activity of the compound<sup>[28]</sup>. The gap between the LUMO and HOMO energies, called the band gap, illustrates the reactivity of a molecule. A wide gap disfavors the excitation of an electron from the HOMO to the LUMO, which consequently leads to a weaker affinity of the inhibitor to the target protein. Therefore, compounds with smaller band gaps are comparatively more reactive<sup>[29]</sup>. In this study, the HOMO and LUMO energies were calculated using four active, two moderately active, and one inactive compound from the training set as well as the 30 hit compounds obtained from the molecular docking studies. The orbital energies of the hit compounds and the known inhibitors were compared to analyze the energy transfer and stability of small molecules in the active site of the protein. An inverse correlation between the HOMO energy and the IC<sub>50</sub> values of known AKR1B10 inhibitors indicated



that the HOMO energy may promote the transfer of electrons from the HOMO of the inhibitor to residues in the active site of AKR1B10. A higher energy value of the HOMO in a compound corresponds to greater activity<sup>[19, 28]</sup>. Only five hit compounds showed higher HOMO values than the already known inhibitors (Table 4). Furthermore, only three compounds were selected based on their lowest band gap energies. The moderately active and inactive compounds exhibited high energy gaps that were not suitable for the reactivity of the molecules. Our results were consistent with published research in the literature<sup>[19, 28, 29]</sup>. Thus, the DFT results indicated that the hit compounds have equivalent or greater electronic properties than most of the active compounds. Finally, three hit compounds that mapped well to the pharmacophoric features of Hypo 1 (Figure 7) were characterized as final hits.

### Molecular dynamics simulations

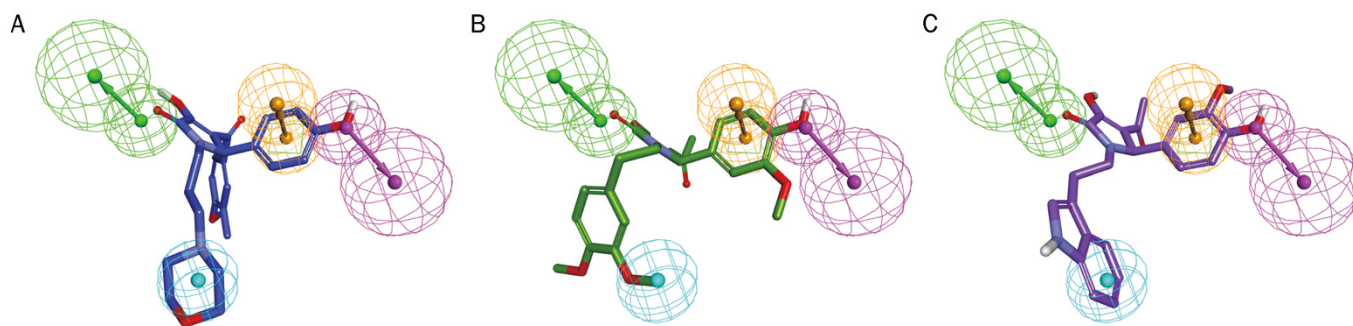
MD simulations were used to assess the binding stability of the final hit-AKR1B10 complexes. All four systems (Table 5) were subjected to 10 ns MD simulations. Thereafter, the RMSD values of the protein backbone atoms and the hit compounds were calculated to observe the overall stability for the protein and ligand. The RMSD values ranged from 0.90 Å to 1.30 Å for the proteins and 0.90 Å to 1.80 Å for all ligands (Figure 8A and 8B). The average RMSD values obtained during the 10 ns simulation time were 1.36 Å, 0.90 Å, 0.90 Å, and 1.85 Å for the inhibitor, hit 1, hit 2, and hit 3, respectively. Variations in the RMSD values of the ligands during the initial 2 ns of the MD simulations were due to the initial adjustment

of the ligand into the active site of AKR1B10. However, the ligands showed more stability in the binding site of AKR1B10 after 2 ns, and no significant changes were observed in protein structure (Figure 8A). The binding modes of the hit compounds were analyzed using their representative structures taken from the last 2 ns. All structures were superimposed and it was found that the binding patterns of the hit compounds were similar to the binding pattern of the reference

**Table 4.** Comparison of orbital energy values of hit compounds and training set compounds calculated by DFT.

| Name                  | HOMO (eV) | LUMO (eV) | $\Delta E^a$ (eV) | IC <sub>50</sub> (nmol/L) |
|-----------------------|-----------|-----------|-------------------|---------------------------|
| Asinex_139155 (Hit 1) | -0.189    | -0.067    | 0.122             |                           |
| Asinex_178779 (Hit 3) | -0.195    | -0.067    | 0.127             |                           |
| Asinex_22867          | -0.195    | -0.039    | 0.155             |                           |
| Asinex_132856 (Hit 2) | -0.195    | -0.068    | 0.127             |                           |
| Asinex_142787         | -0.196    | -0.064    | 0.132             |                           |
| NCI_127946            | -0.196    | -0.047    | 0.148             |                           |
| Training 1            | -0.196    | -0.068    | 0.128             | 4.7                       |
| Training 3            | -0.205    | -0.077    | 0.128             | 7.6                       |
| Training 4            | -0.209    | -0.073    | 0.135             | 8.6                       |
| Training 12           | -0.215    | -0.066    | 0.149             | 480                       |
| Training 15           | -0.224    | -0.068    | 0.155             | 1400                      |
| Training 19           | -0.229    | -0.040    | 0.189             | 9600                      |
| Training 23           | -0.244    | -0.042    | 0.202             | 63000                     |

<sup>a</sup> $\Delta E$ : energy gap (band gap) is the difference between HOMO and LUMO.

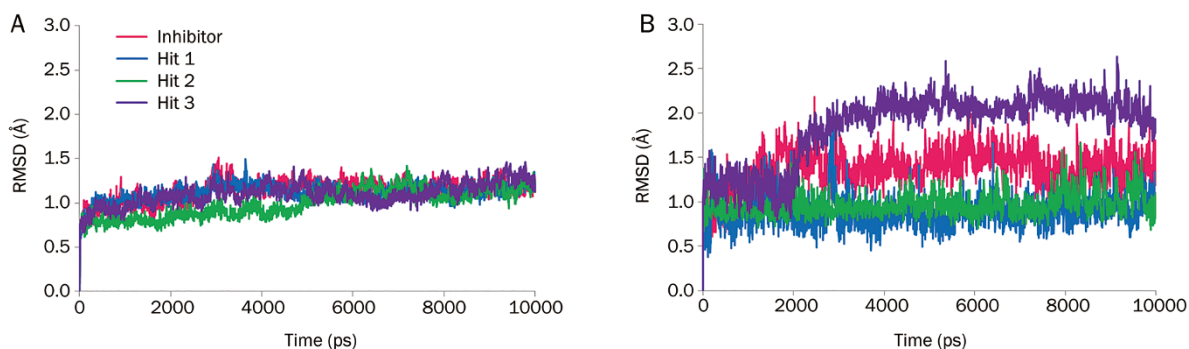


**Figure 7.** Final hit compounds mapped to the best pharmacophore model, Hypo 1. (A) Hit 1; (B) Hit 2; (C) Hit 3. The HBA, HBD, RA, and HYP features are displayed in green, magenta, orange, and cyan, respectively. Hit compounds are represented as stick models.

**Table 5.** The specifications of four systems used for molecular dynamics simulations.

| No | System                               | No of TIP3P water molecules | No of Na <sup>+</sup> counter ions | System size (nm)  |
|----|--------------------------------------|-----------------------------|------------------------------------|-------------------|
| 1  | AKR1B10+NADPH+inhibitor <sup>a</sup> | 14275                       | 4                                  | 5.574×4.725×5.815 |
| 2  | AKR1B10+NADPH+hit 1                  | 14270                       | 4                                  | 5.574×4.725×5.815 |
| 3  | AKR1B10+NADPH+hit 2                  | 14271                       | 4                                  | 5.574×4.725×5.815 |
| 4  | AKR1B10+NADPH+hit 3                  | 14271                       | 4                                  | 5.574×4.725×5.815 |

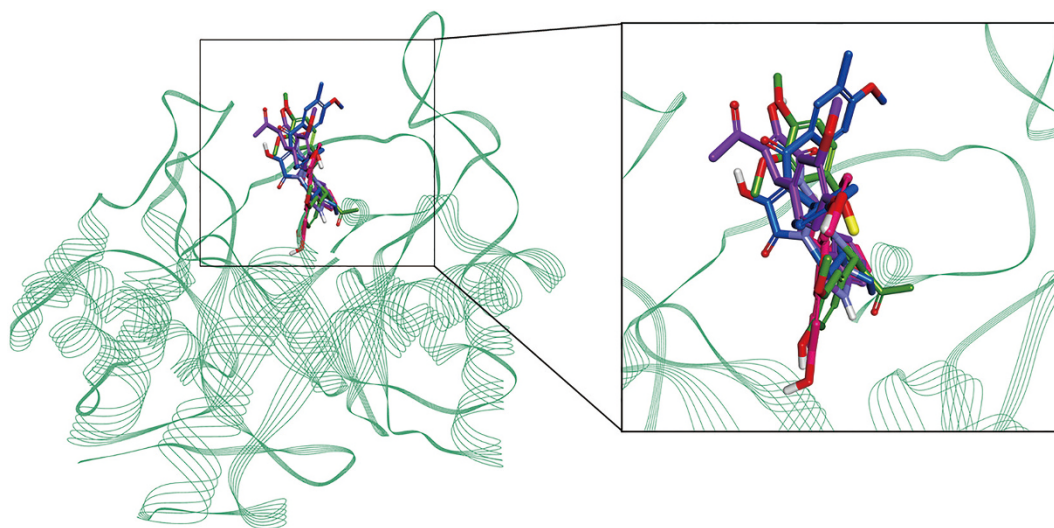
<sup>a</sup>Reference inhibitor; the most active compound in the training set, 7-hydroxy-2-(4-methoxyphenylimino)-2H-chromene-3-carboxylic acid benzylamide.



**Figure 8.** RMSD plots for checking the overall stability of the systems during 10 ns MD simulations. (A) The RMSD profile for the backbone atoms of the AKR1B10 protein. (B) RMSD profile for bound compounds. The colored lines represent the AKR1B10 protein with a reference inhibitor (magenta), Hit 1 (blue), Hit 2 (green), and Hit 3 (purple).

inhibitor (Figure 9). The inhibitor and hit compounds were positioned between the external loops B and C. The substrate binding pocket in AKR1B10 is defined by Tyr49, His111, and Trp112. These important residues were also found in the binding of the reference inhibitor and the hit compounds (Table 6). The 7-hydroxyl group of the reference inhibitor formed a hydrogen bond with the important residue Tyr49 in the active site of AKR1B10 (Figure 10A) and showed hydrogen bond interactions with Lys22, Lys78 and a weak hydrogen bond with His111 (3.356 Å, not shown in the figure). Apart from hydrogen bond formation, the inhibitor showed interactions with hydrophobic pocket residues such as Val48, Trp112, Phe123, Cys299, and Val301, which may have imparted its specificity as an AKR1B10 inhibitor<sup>[17]</sup>. Hit 1 formed hydrogen bonds with Tyr49, Lys22, and Gln50 (Figure 10B). More interestingly, hit 1 also formed a hydrogen bond with Lys125, which is present at the entrance to the binding pocket and is important for the specific binding to AKR1B10. Hit 1 showed

hydrophobic interactions with Val48, Trp80, His111, Trp112, Phe123, Asn161, Pro219, Trp220, Cys299, Val301, and Leu302. Hit 2 formed hydrogen bonds with Tyr49, Cys299, and Val301 as well as weak hydrogen bonds with Lys78 (3.446 Å) and His111 (3.399 Å) (Figure 10C). Hit 2 showed interactions with hydrophobic pocket residues such as Trp21, Val48, Trp80, His111, Trp112, Phe123, Trp220 and others, including Ala46, Lys78, Asn161, and Leu302. Hit 3 showed a slightly different binding mode, forming a hydrogen bond with His111 and a weak hydrogen bond with Tyr49 (3.293 Å) (Figure 10D). Hit 3 also formed hydrogen bonds with Val48 and Gln50 and showed hydrophobic interactions with specificity pocket residues, including Trp21, Tyr49, Trp80, Trp112, Phe123, Pro219, Trp220, Val301, and Leu302. The interaction details of the interacting residues are summarized in Table 6. The binding mode analysis for the hit compounds revealed that they bind to the active site via molecular interactions such as hydrogen bonds and hydrophobic interactions. Moreover, the hit

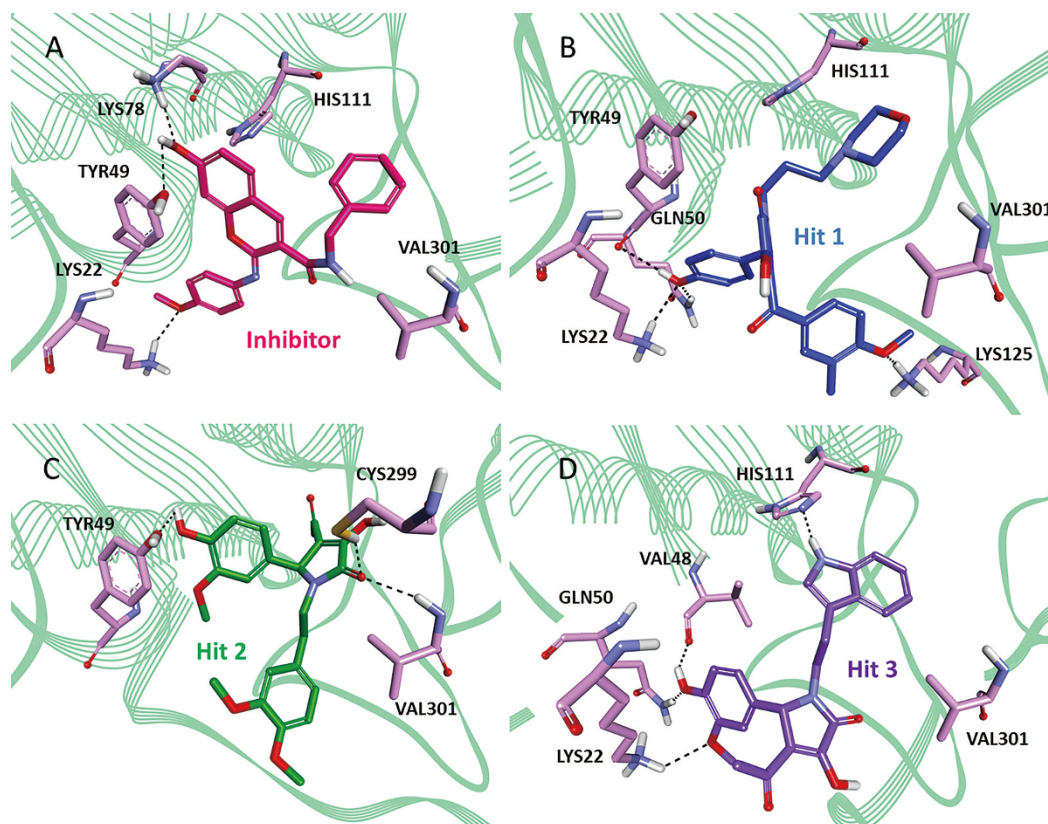


**Figure 9.** The binding patterns of the reference inhibitor and three hit compounds in the active site of AKR1B10. All compounds in their representative structures were superimposed (left) and enlarged (right). The protein is shown in green color. Magenta, blue, green and purple represent the AKR1B10 protein with the reference inhibitor, Hit 1, Hit 2, and Hit 3, respectively. Only polar hydrogen atoms are shown for clear visualization.

**Table 6.** The molecular interactions between the compounds and AKR1B10 protein.

| Compound               | Hydrogen bond interaction |            |                 |            | Hydrophobic interaction  |
|------------------------|---------------------------|------------|-----------------|------------|--|
|                        | Ligand atom               | Amino acid | Amino acid atom | Length (Å) |  |
| Inhibitor <sup>a</sup> | O23                       | Lys22      | HZ3             | 2.336      | Val48, Gln50, His111, Trp112, Phe123, Cys299, Val301, Leu302                         |
|                        | O24                       | Tyr49      | OH              | 2.446      |  |
|                        | O24                       | Lys78      | HZ2             | 1.939      |  |
| Hit 1                  | O28                       | Lys22      | HZ3             | 2.317      | Val48, Trp80, His111, Trp112, Phe123, Asn161, Pro219, Trp220, Cys299, Val301, Leu302 |
|                        | O28                       | Gln50      | HE21            | 2.051      |  |
|                        | H50                       | Tyr49      | O               | 2.330      |  |
|                        | O27                       | Lys125     | HZ1             | 1.988      |  |
| Hit 2                  | H44                       | Tyr49      | OH              | 2.442      | Trp21, Ala46, Val48, Lys78, Trp80, His111, Trp112, Phe123, Asn161, Trp220, Leu302    |
|                        | O11                       | Cys299     | HG              | 2.021      |  |
|                        | O11                       | Val301     | HN              | 2.494      |  |
| Hit 3                  | O23                       | Lys22      | HZ2             | 2.482      | Trp21, Tyr49, Trp80, Trp112, Phe123, Pro219, Trp220, Val301, Leu302                  |
|                        | H42                       | Val48      | O               | 1.474      |  |
|                        | HE21                      | Gln50      | O24             | 2.051      |  |
|                        | H32                       | His111     | NE2             | 1.676      |  |

<sup>a</sup>Reference inhibitor; the most active compound in the training set, 7-hydroxy-2-(4-methoxyphenylimino)-2*H*-chromene-3-carboxylic acid benzylamide.

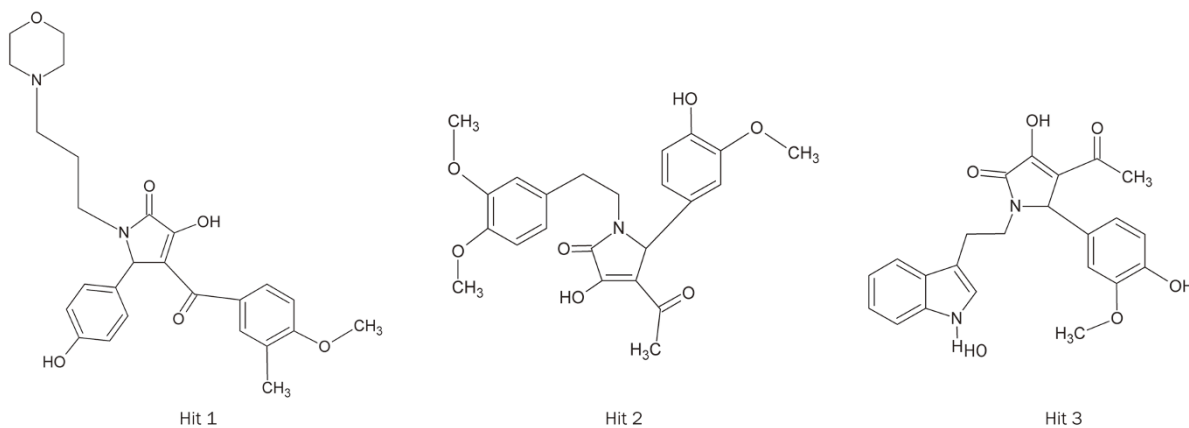


**Figure 10.** The binding mode analysis of the reference inhibitor and the three hit compounds in the active site of AKR1B10. (A) Inhibitor: magenta; (B) Hit 1: blue; (C) Hit 2: green; and (D) Hit 3: purple. The hydrogen bond interactions between compounds and protein residues (pink color) are shown as black dotted lines. Only polar hydrogen atoms are shown for clear visualization.

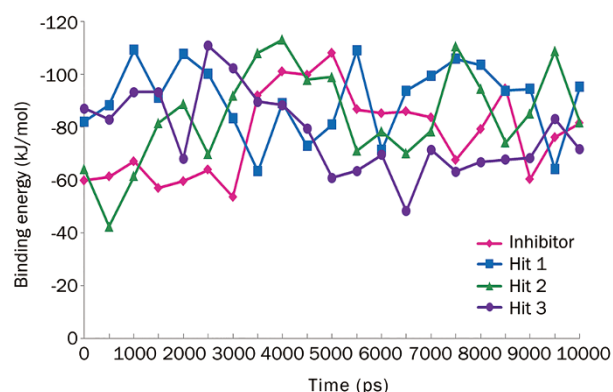
compounds showed a large number of hydrophobic interactions with the residues that determine specificity for selective inhibition of AKR1B10. The novelty of hit compounds was checked by a similarity search using the online search tool *Pubchem Structure*<sup>[45]</sup>. The search results confirmed that the hit compounds have not been previously tested experimentally for the inhibition of AKR1B10. These results conclude that the final hit compounds can be recommended as potential new scaffolds for designing potent and selective inhibitors of AKR1B10. The 2D structures of hit compounds are shown in Figure 11.

#### Analysis of the binding free energy of AKR1B10 and reference inhibitor/hit compounds

The binding free energy calculations after MD simulation is a time-demanding step, as reported earlier<sup>[46, 47]</sup>. Logically, the  $\Delta G$  values obtained from the sets of snapshot structures generated during the MD simulation should consider the fluctuation of the protein and the ligand conformation in the complex, thereby ensuring an appropriate adjustment of the ligand in the binding site<sup>[42, 46]</sup>. The MM/PBSA calculations of the AKR1B10-ligand complexes using the reference inhibitor, hit 1, hit 2, and hit 3 as the ligands produced favorable  $\Delta G$  values in the range of -45 to -115 kJ/mol, as depicted in Figure 12. The binding energy showed slight variation in each snapshot because the conformational space was not sampled enough to obtain converged results. The average binding energy gained for the AKR1B10-ligand complexes were -77.2 kJ/mol (reference inhibitor as ligand), -90.3 kJ/mol (hit 1), -84.2 kJ/mol (hit 2), and -77.5 kJ/mol (hit 3) (Figure 12, Table 7). Additionally, the binding free energy calculations of the hit compounds were compared with the binding affinity assessed by the docking experiments (Table 7). The hit compounds represented higher Goldscores and lower binding free energies than the reference inhibitor. These results demonstrate that the hit compounds have stronger binding affinities than the reference inhibitor.



**Figure 11.** 2D structures of the final three hit compounds. All three hit compounds were identified from ASINEX database.



**Figure 12.** Graphical representation of MM/PBSA estimated binding free energy of AKR1B10 and reference inhibitor/hit 1/hit 2/hit 3 complex throughout simulation time. Color coding: Inhibitor: magenta, Hit 1: blue, Hit 2: green, and Hit 3: purple.

**Table 7.** Comparison of average binding energy and Goldscore of AKR1B10 and reference inhibitor/hit 1/hit 2/hit 3 complex.

| Systems                        | Average binding energy (kJ/mol) | Gold fitness score |
|--------------------------------|---------------------------------|--------------------|
| AKR1B10+inhibitor <sup>a</sup> | -77.2                           | 37.8               |
| AKR1B10+hit 1                  | -90.3                           | 68.5               |
| AKR1B10+hit 2                  | -84.2                           | 61.0               |
| AKR1B10+hit 3                  | -77.5                           | 54.0               |

<sup>a</sup>Reference inhibitor; the most active compound in the training set, 7-hydroxy-2-(4-methoxyphenylimino)-2H-chromene-3-carboxylic acid benzylamide.

#### Discussion

In the past five years, AKR1B10 has emerged as a highly rated tumor marker in various malignancies, including smoker's

breast cancer, uterine cancer, pancreatic cancer, lung cancer, and colorectal cancer. Therefore, AKR1B10 is a potential therapeutic target for the treatment of various human cancers. AKR1B10 shares up to 70% structural similarity with AKR1B1 a member of the same family of proteins. Because various AKR1B1 inhibitors (*eg*, tolrestat, epalrestat, sorbinil and fidarostat) can also inhibit AKR1B10 due to the high structural similarity of these two proteins, inhibitors that can selectively inhibit AKR1B10 are of primary importance. Thus, a ligand-based pharmacophore model approach combined with other computational techniques was employed in the design of novel and selective AKR1B10 inhibitors. The best pharmacophore model (Hypo 1) containing one HBA, one HBD, one RA, and one HYP feature was built. The model was found to be the best representative model as it showed a high correlation value, the lowest total cost and the least RMSD value. Hypo 1 was validated using Fischer's randomization and test set methods. A high correlation between the experimental and predicted IC<sub>50</sub> values in the training and test sets indicated a good predictive ability for Hypo 1. Hypo 1 was used as a 3D query for database screening and a total of 125 compounds were screened by applying Lipinski's Rule of Five and ADMET calculations. These compounds were docked into the active site of AKR1B10 to analyze their binding conformations. The compounds with high fitness scores that showed interactions with important residues of AKR1B10 were selected for further design as AKR1B10 inhibitors. Consequently, 30 compounds were assessed using molecular orbital energy (HOMO and LUMO) calculations to confirm the reactivities of the final hits. Only three compounds represented higher HOMO values and small band gaps when compared to the training set compounds. The reasonable binding modes of the hit compounds were assessed using 10 ns MD simulations. All hit compounds presented similar binding modes in the active site of AKR1B10 as the reference inhibitor. Furthermore, all of the hit compounds showed a greater extent of hydrophobic interactions with residues in the specificity pocket, which may be involved in their specific inhibitory activities against AKR1B10. The MM-PBSA analyses of the binding energies showed lower energies for the AKR1B10-ligand complexes using hit 1, hit 2, and hit 3 as the ligand and the highest binding affinities due to the presence of hydrogen bonding and hydrophobic interactions in each of these complexes. The free energy calculations using MM/PBSA method have shown to produce reliable and reproducible predictions of binding affinity that could be used in drug discovery and personalized medicine applications<sup>[48]</sup>. It can be helpful to rank the lead compounds according to their predicted affinities to the target and this information may be used to guide molecular design and synthesis<sup>[49]</sup>. However, it will be interesting to assess the experimental inhibitory activities of the hit compounds against AKR1B10. These results indicate that the final hit compounds are promising candidates for use as selective AKR1B10 inhibitors.

In conclusion, by using a set of known active AKR1B10 inhibitors, we have generated a pharmacophore model that

can be used to map active compounds. The compounds obtained from pharmacophore mapping were subjected to docking studies with the active site of AKR1B10. Molecular orbital energy calculations provided further insights into the reactivities of the hit compounds. MD simulation analyses were conducted to identify the binding modes of hit compounds. The results of these analyses suggested that the hit compounds bind to AKR1B10 in a similar manner as the reference inhibitor. The binding free energy calculations revealed that the hit compounds had higher binding affinities towards AKR1B10 than the reference inhibitor. We believe that these new scaffolds, which include the final hits of this study, will be helpful in designing novel drugs for AKR1B10-related diseases.

### Acknowledgements

This research was supported by the Basic Science Research Program (2012R1A1A4A01013657) and Management of Climate Change Program (2010-0029084) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education of Republic of Korea. This work was also supported by the Next-Generation BioGreen 21 Program (PJ009486) from the Rural Development Administration (RDA) of Republic of Korea.

### Author contribution

Raj KUMAR, Keun Woo LEE, Minky SON, and Rohit BAVI designed the research; Raj KUMAR and Rohit BAVI performed the experiments; Keun Woo LEE, Yuno LEE, Hyong-ha KIM, Jung-keun SUH, Yong-seong KIM, and Yong Jung KWON analyzed the results; and Raj KUMAR, Chanin PARK, Venkatesh ARULALAPPERUMAL and Guang Ping CAO prepared the manuscript.

### Supplementary information

Supplementary information is available at *Acta Pharmacologica Sinica's* website.

### References

- 1 Gallego O, Ruiz FX, Ardevol A, Dominguez M, Alvarez R, de Lera AR, *et al*. Structural basis for the high all-*trans*-retinaldehyde reductase activity of the tumor marker AKR1B10. *Proc Natl Acad Sci U S A* 2007; 104: 20764–9.
- 2 Chung YT, Matkowskyj KA, Li H, Bai H, Zhang W, Tsao MS, *et al*. Overexpression and oncogenic function of aldo-keto reductase family 1B10 (AKR1B10) in pancreatic carcinoma. *Mod Pathol* 2012; 25: 758–66.
- 3 Cao D, Liao DF. Author's reply to: AKR1B10 and its emerging role in tumor carcinogenesis and as a cancer biomarker. *Int J Cancer* 2013; 132: 496–7.
- 4 Kapoor S. AKR1B10 and its emerging role in tumor carcinogenesis and as a cancer biomarker. *Int J Cancer* 2013; 132: 495.
- 5 Fukumoto S, Yamauchi N, Moriguchi H, Hippo Y, Watanabe A, Shibahara J, *et al*. Overexpression of the aldo-keto reductase family protein AKR1B10 is highly correlated with smokers' non-small cell lung carcinomas. *Clin Cancer Res* 2005; 11: 1776–85.
- 6 Zhang W, Li H, Yang Y, Liao J, Yang GY. Knockdown or inhibition of

- aldo-keto reductase 1B10 inhibits pancreatic carcinoma growth via modulating Kras-E-cadherin pathway. *Cancer Lett* 2014; 355: 273–80.
- 7 Ruiz FX, Gallego O, Ardevol A, Moro A, Dominguez M, Alvarez S, et al. Aldo-keto reductases from the AKR1B subfamily: retinoid specificity and control of cellular retinoic acid levels. *Chem Biol Interact* 2009; 178: 171–7.
  - 8 Balendiran GK, Martin HJ, El-Hawari Y, Maser E. Cancer biomarker AKR1B10 and carbonyl metabolism. *Chem Biol Interact* 2009; 178: 134–7.
  - 9 Matsunaga T, Yamane Y, Iida K, Endo S, Banno Y, El-Kabbani O, et al. Involvement of the aldo-keto reductase, AKR1B10, in mitomycin-c resistance through reactive oxygen species-dependent mechanisms. *Anticancer Drugs* 2011; 22: 402–8.
  - 10 Zhang L, Zhang H, Zhao Y, Li Z, Chen S, Zhai J, et al. Inhibitor selectivity between aldo-keto reductase superfamily members AKR1B10 and AKR1B1: role of Trp112 (Trp111). *FEBS Lett* 2013; 587: 3681–6.
  - 11 Wang C, Yan R, Luo D, Watabe K, Liao DF, Cao D. Aldo-keto reductase family 1 member B10 promotes cell survival by regulating lipid synthesis and eliminating carbonyls. *J Biol Chem* 2009; 284: 26742–8.
  - 12 Yan R, Zu X, Ma J, Liu Z, Adeyanju M, Cao D. Aldo-keto reductase family 1 B10 gene silencing results in growth inhibition of colorectal cancer cells: Implication for cancer intervention. *Int J Cancer* 2007; 121: 2301–6.
  - 13 Matkowskyj KA, Bai H, Liao J, Zhang W, Li H, Rao S, et al. Aldoketo-reductase family 1B10 (AKR1B10) as a biomarker to distinguish hepatocellular carcinoma from benign liver lesions. *Hum Pathol* 2014; 45: 834–43.
  - 14 Soda M, Hu D, Endo S, Takemura M, Li J, Wada R, et al. Design, synthesis and evaluation of caffeic acid phenethyl ester-based inhibitors targeting a selectivity pocket in the active site of human aldo-keto reductase 1B10. *Eur J Med Chem* 2012; 48: 321–9.
  - 15 Takemura M, Endo S, Matsunaga T, Soda M, Zhao HT, El-Kabbani O, et al. Selective inhibition of the tumor marker aldo-keto reductase family member 1B10 by oleanolic acid. *J Nat Prod* 2011; 74: 1201–6.
  - 16 Endo S, Matsunaga T, Kuwata K, Zhao HT, El-Kabbani O, Kitade Y, et al. Chromene-3-carboxamide derivatives discovered from virtual screening as potent inhibitors of the tumour maker, AKR1B10. *Bioorg Med Chem* 2010; 18: 2485–90.
  - 17 Endo S, Hu D, Suyama M, Matsunaga T, Sugimoto K, Matsuya Y, et al. Synthesis and structure-activity relationship of 2-phenylimino-chromene derivatives as inhibitors for aldo-keto reductase (AKR) 1B10. *Bioorg Med Chem* 2013; 21: 6378–84.
  - 18 Niu MM, Qin JY, Tian CP, Yan XF, Dong FG, Cheng ZQ, et al. Tubulin inhibitors: pharmacophore modeling, virtual screening and molecular docking. *Acta Pharmacol Sin* 2014; 35: 967–79.
  - 19 Sakkiah S, Lee KW. Pharmacophore-based virtual screening and density functional theory approach to identifying novel butyryl-cholinesterase inhibitors. *Acta Pharmacol Sin* 2012; 33: 964–78.
  - 20 Li H, Sutter J, Hoffmann R. HypoGen: An automated system for generating predictive 3D pharmacophore models. In: Güner OF, editor. *Pharmacophore Perception, Development, and use in Drug Design*. California: International University Line; 2000. p 171–189.
  - 21 Sakkiah S, Thangapandian S, John S, Kwon YJ, Lee KW. 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. *Eur J Med Chem* 2010; 45: 2132–40.
  - 22 Debnath AK. Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of mycobacterium avium complex dihydrofolate reductase. *J Med Chem* 2002; 45: 41–53.
  - 23 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001; 46: 3–26.
  - 24 Yang SY. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 2010; 15: 444–50.
  - 25 Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997; 267: 727–48.
  - 26 Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins* 2003; 52: 609–23.
  - 27 Tsuneda T. Orbital energy. In: Tsuneda T, editor. *Density functional theory in quantum chemistry*. Japan: Springer; 2014. p 161–188
  - 28 Eroglu E, Turkmen H. A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II. *J Mol Graph Model* 2007; 26: 701–8.
  - 29 Ai C, Li Y, Wang Y, Li W, Dong P, Ge G, et al. Investigation of binding features: effects on the interaction between CYP2A6 and inhibitors. *J Comput Chem* 2010; 31: 1822–31.
  - 30 Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* 2005; 26: 1701–18.
  - 31 Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem* 2011; 32: 2359–68.
  - 32 Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys* 2007; 126: 014101.
  - 33 Parrinello M, Rahman A. Polymorphic transitions in single crystals. A new molecular dynamics method. *J Appl Phys* 1981; 52: 7182.
  - 34 Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997; 18: 1463–72.
  - 35 Darden T, York D, Pedersen L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J Chem Phys* 1993; 98: 10089.
  - 36 Son M, Baek A, Sakkiah S, Park C, John S, Lee KW. Exploration of virtual candidates for human HMG-CoA reductase inhibitors using pharmacophore modeling and molecular dynamics simulations. *PLoS One* 2013; 8: e83496.
  - 37 Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996; 14: 33–8.
  - 38 Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 2011; 51: 69–82.
  - 39 Spiliotopoulos D, Spitaleri A, Musco G. Exploring PHD fingers and H3K4meO interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS One* 2012; 7: e46902.
  - 40 Sonawane KD, Barage SH. Structural analysis of membrane-bound hECE-1 dimer using molecular modeling techniques: insights into conformational changes and A $\beta_{1-42}$  peptide binding. *Amino Acids* 2014. doi: 10.1007/s00726-014-1887-8 Epub 2014 Dec 16.
  - 41 Genheden S, Ryde U. How to obtain statistically converged MM/GBSA results. *J Comput Chem* 2010; 31: 837–46.
  - 42 Vorontsov II, Miyashita O. Crystal molecular dynamics simulations to speed up MM/PB(GB)SA evaluation of binding free energies of dimannose deoxy analogs with P51G-m4-Cyanovirin-N. *J Comput Chem* 2011; 32: 1043–53.

- 43 Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997; 11: 425–45.
- 44 Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, Mortenson PN, *et al*. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem* 2007; 50: 726–41.
- 45 Wang Y, Suzek T, Zhang J, Wang J, He S, Cheng T, *et al*. PubChem BioAssay: 2014 update. *Nucleic Acids Res* 2014; 42 (Database issue): D1075–82.
- 46 Bonnet P, Bryce RA. Molecular dynamics and free energy analysis of neuraminidase–ligand interactions. *Protein Sci* 2004; 13: 946–57.
- 47 Donini OA, Kollman PA. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J Med Chem* 2000; 43: 4180–8.
- 48 Wright DW, Hall BA, Kenway OA, Jha S, Coveney PV. Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *J Chem Theory Comput* 2014; 10:1228–41.
- 49 Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* 2007; 36: 21–42.