JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Discovery of New Selective Human Aldose Reductase Inhibitors through Virtual Screening Multiple Binding Pocket Conformations

Ling Wang, Qiong Gu, Xuehua Zheng, Jiming Ye, Zhihong Liu, Jiabo Li, Xiaopeng Hu, Arnold T. Hagler, and Jun Xu

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/ci400322j • Publication Date (Web): 31 Jul 2013

Discovery of New Selective Human Aldose Reductase Inhibitors through Virtual Screening Multiple Binding Pocket Conformations

Ling Wang,[†] Qiong Gu, [†]* Xuehua Zheng,[†] Jiming Ye,[‡] Zhihong Liu,[†] Jiabo Li,[§] Xiaopeng Hu,[†] Arnold Hagler,[#] and Jun Xu[†]*

[†]*Research Center for Drug Discovery & Institute of Human Virology, School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510006, China*

[#]*Department of Chemistry, University of Massachusetts, 701 Lederle Graduate Research Tower, 710 North Pleasant Street, Amherst, Massachusetts 01003-9336, United States; Shifa Biomedical, 1 Great Valley Parkway, Suite 8, Malvern, Pennsylvania 19355, United States*

[‡] *Molecular Pharmacology for Diabetes Group, Health Innovations Research Institute and School of Health Sciences, RMIT University, Melbourne, VIC 3083, Australia*

[§]*Accelrys Inc., 10188 Telesis Court, San Diego, California 92121*

ABSTRACT

Aldose reductase reduces glucose to sorbitol. It plays a key role in many of the complications arising from diabetes. Thus, Aldose reductase inhibitors (ARI) have been identified as promising therapeutic agents for treating such complications of diabetes, as neuropathy, nephropathy, retinopathy, and cataracts. In this paper, a virtual screening protocol applied to a library of compounds in house, has been utilized to discover novel ARIs. IC_{50}'s were determined for 15 hits that inhibited ALR2 to greater than 50% at 50 μM, and ten of these have an IC_{50} of 10 μM or less, corresponding to a rather substantial hit rate of 14% at this level. The specificity of these compounds relative to their cross-reactivity with human ALR1 was also assessed by inhibition assays. This resulted in identification of novel inhibitors with IC_{50}'s comparable to the commercially available drug, epalrestat and greater than an order of magnitude better selectivity.

INTRODUCTION

Diabetes patients suffer from long-term complications, such as neuropathy, nephropathy, retinopathy and cataracts.[1, 2] Although the mechanisms of diabetic complications are not completely understood, many biochemical pathways associated with hyperglycemia have been identified.[3] The sorbitol pathway is one of the most important pathways implicated in long-term complications.[3] Aldose reductase enzyme (ALR2, EC1.1.1.21, AKR1B1) is a member of the aldo-ketoreductase (AKR) superfamily, and, together with sorbitol dehydrogenase forms the polyol pathway.[4] ALR2 is the rate-limiting enzyme in this pathway. It reduces the aldehyde form of glucose to sorbitol by using NADPH as a cofactor. Then, sorbitol dehydrogenase converts the sorbitol to fructose by using NAD^+ as a cofactor.[5] Under normal circumstances, the affinity of ALR2 and glucose is low. While, under hyperglycemic circumstances, highly expressed ALR2 results in twofold to fourfold accelerated conversion of glucose to sorbitol. However, the rate of sorbitol dehydrogenase metabolism is not affected, which results in significant sorbitol accumulation under hyperglycemic circumstances. The sorbitol accumulation leads to osmotic swelling, changes in membrane permeability, and oxidative stress culminating in tissue injury associated with late-onset diabetic complications.[6]

According to recent reports, ALR2 is up-regulated not only under hyperglycemic conditions, but also in other pathological states including cardiac disorders (myocardial ischemia and ischemia-reperfusion injury, congestive heart failure, cardiac hypertrophy, and cardiomyopathy), inflammation, mood disorders, renal

insufficiency, ovarian abnormalities, and human cancers such as liver, breast, ovarian, cervical, and rectal cancers.[7-9] Those pathological processes have become major threats to human health in the 21st century.

Because of these observations, aldose reductase has emerged as an attractive therapeutic target for long-term diabetic complications, cardiac disorders, and inflammatory diseases. Intense efforts have been directed toward the development of effective aldose reductase inhibitors,[10] however, only a few compounds have reached clinical trials, such as, alrestatin,[11] tolrestat,[12] epalrestat,[13] zopolrestat,[14] zenarestat,[15] ponalrestat,[16] and lidorestat[17]. So far, epalrestat (Kinedak), marketed in Japan and China, is the only commercially available ARI drug. Aside from epalrestat ALR2 inhibitors have failed in clinical trials because of poor pharmacokinetic properties and side effects[18] and even epalrestat has been withdrawn from the market in other countries because of its side effects. Thus it is important to develop novel ARIs with improved efficacy, pharmacodynamics, pharmacokinetic properties and safety profile.

X-ray studies reveal that there are at least three distinct binding pocket conformations of ALR2, corresponding to three different ligand types.[19] These binding pockets were reported by Sandro and coworkers (PDB codes: 2PDK[20], 1US0,[21] and 2FZD[22,23]). Comparative structural analysis and molecular dynamic (MD) simulations studies indicate that for ALR2, a single experimental structure is not sufficient to predict all possible binding modes;[19] and a higher virtual screening score does not necessarily correspond to higher biological activity because of false negatives from the docking procedure.[10, 24] These deficiencies result in lower virtual

screening hit rates.

Therefore, we exploited a virtual screening protocol, combined with MD simulations to overcome some of these issues. Starting with three experimental structures (PDB entries: 2PDK, 1US0 and 2FZD), we use MD simulations to sample accessible binding site conformers around each. The final binding site conformations are then derived by averaging the conformers from the three MD simulation results, respectively. The compound library is then virtually screened against these three averaged structures, and the docked complexes were optimized by MD simulations to assess their stability. As MD simulations are extremely computationally demanding and in general intractable to apply to numerous ligand-protein systems ligands, we exploited Graphics Processing Unit (GPU) technology, which significantly accelerates the calculation relative to more conventional central processing units (CPUs).[25, 26] The compounds selected through this virtual screening protocol were tested for ALR2 inhibition in vitro, and highly active ARIs containing new chemotypes were identified. In addition, the selectivity of compounds demonstrating potent ALR2 inhibition was assessed by testing for ALR1 inhibition, and their toxicity was also tested by MTT (3-(4,5-dimethylthiazol-2-yl)-2-5 diphenyltetrazolium bromide) assays: they showed considerable selectivity and no evidence of cell toxicity.

RESULTS AND DISCUSSION

***In Silico* Screening Against ALR2.** Accounting for protein flexibility and induced fit effects continues to be a challenge in virtual screening efforts.[27, 28] This is

especially true for the human ALR2 binding pocket.[19, 29, 30] The significant mutual induced-fit effects upon binding different ligands to ALR2, provides a challenge for standard structure-based virtual screening.[31-33] Virtual screening can be done via either docking molecules into a single binding pocket conformation derived from an experimental structure,[10, 34, 35] or docking molecules into multiple binding pocket conformations derived from multiple experimental structures simultaneously.[36] For example in previous work, using a clustering approach of diverse binding site volumetric shapes, we chose four representative structures as a compromise between more extensive sampling and computational tractability.[27, 28] Cosconati and coworkers reported that the multiple binding pocket strategy applied to ALR2 resulted in a higher virtual screening hit rate.[23] Given the significant plasticity of this protein, the question arises as to whether a more extensive sampling of ALR2 conformational space could further improve either enrichment or diversity of ligands recovered.[19] The protocol we use here to further account for ligand-receptor flexibility is given in Figure 1. The virtual screening protocol starts with the three ALR2-ligand complexes (PDB codes: 2PDK, 1US0, and 2FZD) that represent three types of static binding modes. MD simulations were applied to these structures for six nanoseconds (ns) to refine the protein structure especially for regions that may be poorly resolved in the X-ray. The time averaged structures over the last nanosecond of the trajectories (See Supporting Information, Figure S1) were then, minimized by steepest descent (Discovery Studio 2.1, Accelrys Inc., San Diego, CA.) to refine covalent geometry and remove collisions. These refined structures were then employed for docking in

the first stage of the virtual screening campaign.

The GSMTL (Guangdong Small Molecule Tangible Library), a small molecule repository containing more than 7,200 compounds, was selected for screening against ALR2.[37] The compounds in the library were docked against the three refined binding pocket conformations with FlexX. In order to establish criteria for the virtual screening hits, we have: (1) docked 927 known ALR2 inhibitors[38] into the three binding pockets to calibrate the relation between activity and FlexX score, generating an averaged active compound FlexX docking score (ACFDS), and (2) computed active ALR2 inhibitor protein ligand interaction fingerprints (PLIF) from 76 co-crystal structures in MOE (Molecular Operating Environment[39], Figure 2).

The virtual screening hits then have to pass three filters:

(1) The FlexX docking score must be below the ACFDS. 1,238 compound hits passed this filter.

(2) Hits, which have acceptable scores, are retained only if their PLIFs match the active ALR2 inhibitor PLIF (at least two hydrogen bonds with Tyr48, His110, or Trp111, and hydrophobic interaction with the specificity hydrophobic pocket). 128 compound hits passed this filter.

(3) Each of the 128 hits was subjected to MD simulations. Only hits with RMSD (root-mean-square deviation) values less than 3 Å from the docked pose were retained. This is an "internal consistency" check to ensure that the docked conformation scored is indeed the stable conformation of the system. 71 compounds survived this criterion. These final hits were further tested by

bioassays. (Note each compound had three binding modes corresponding to the three ALR2 binding pocket conformations described above (Figure 1); if a compound passed through the first two filters bound to more than one of the structures the pose with the highest docking score was selected for this last MD refinement stage.)

We note that our primary goal is to reduce the number of compounds that need to be subjected to more expensive experimental screening while achieving a good "yield" of a reasonable number of hits i.e. remove false positives. The filters outlined above indubitably remove some true positives as well, as does the inherent virtual screening algorithm itself, but as long as we achieve the requisite number and quality (diversity, scaffolds etc.) we have achieved our goal. This is the basic assumption in any screening campaign. Any of these filters might be omitted if sufficient hits are not obtained. Questions naturally arise about the consequences of omitting any of the filters or adding others and also comparing with other algorithms. Some estimate of the utility of the procedure may be obtained by comparison with a simple similarity search. This is given below. A detailed study of the effect of all the variables/filters in the algorithm as well as comparison with other algorithms is the subject of further study. Here our goal was primarily to find new ARIs containing novel chemotypes.

Confirmation of the Hits from Virtual Screening. The 71 final hits were assayed for ALR2 inhibition as described in methods. The bioassays confirmed that

26 of the compounds showed a minimum of 30% inhibition of ALR2 at 50 μM (Table 1). The 26 confirmed ALR2 inhibitors are depicted in Figure 3. IC_{50}'s were determined for the 15 hits which inhibited ALR2 to greater than 50% at 50 μM. In order to ensure that the IC_{50} determinations are reliable, IC_{50} values of quercetin and epalrestat (Figure S2) were obtained and verified to reproduce previously reported values[40, 41] (28 μM and 0.28 μM). The most active compounds are **14** and **25**; they demonstrate sub-micromolar IC_{50} concentrations (0.22 μM and 0.89 μM), values comparable to the commercially available drug, epalrestat. Other promising compounds were **1**, **18**, **22**, and **24**, which exhibited IC_{50} values < 10 μM.

Comparison of Results with Simple Ligand Based Searches. We have carried out both a 2D similarity search as well as a substructure search in order to confirm that these simpler techniques[42, 43] do not recover the novel actives resulting from the docking protocol discussed above. In the similarity search, a total of 22 ALR2 actives derived from the DUD (A Directory of Useful Decoys; http://dud.docking.org/), along with the reference compounds quercetin and epalrestat, shown in Figure S3 were chosen as the reference structures. The two dimensional structural similarities of the structures in the Guangdong Small Molecule Tangible Library (GSMTL) database, with the 24 queries were calculated based on an atom-center fragment approach.[42, 43] Imposing a similarity threshold of 80%, we retrieved 67 hits from GSMTL database. Among these hits, 51 of them were flavone derivatives arising from similarity with quercetin, and 16 "non-flavones". The preponderance of flavones arises presumably due to their ubiquity in herbs and plants.[44] The non-flavanoid hits are shown in Table

S1 along with the reference compound they arose from.

For a second simple similarity search control we exploited a Markush search.[45] Six Markush search queries were prepared as shown in Figure S4. A total of 421 hits were found. Of these the majority included either a flavone scaffold or benzoic sulfonic groups, while 67 of the 421 represent other diverse scaffolds.[46]

The most liberal, first order "atom center fragment"[42, 43] cutoff was used in these similarity studies in order to retrieve the maximum number of compounds, Despite this none of the actives identified in the docking studies were found by even this liberal definition of similarity.

Structural Characteristics of the Confirmed Hits. Like epalrestat, most of these new ALR2 inhibitors contain carboxylic acid moieties. In particular, the two sub-micromolar compounds, **14** and **25** as well as the four compounds with single digit IC_{50} values (**1**, **18**, **22**, and **24)**, all contain the carboxylic acid moiety. The most favorable docking score for **14** from the three binding pocket conformations was -31.39 kcal/mol (Table S2). The predicted binding mode of compound **14** was stable in MD simulations with a time averaged RMSD of ~1 Å over the 6 nanoseconds trajectory (Figure S5 of Supporting Information).

As seen in Figure 4a, the ligand carboxyl inserts into the anion binding pocket, H-bonding with Tyr48, His110, and Trp111 side chains and engages in an electrostatic interaction with the nicotinamide moiety of the $NADP^+$ cofactor. Additional hydrogen bonds were formed between the acyl oxygen of **14** and Trp111. The naphthalene ring of compound **14** occupies ALR2's specificity pocket forming hydrophobic contacts

with Trp79, Trp111, Phe122, Phe115 and Leu300. Not surprisingly these interactions are consistent with the PLIFs, which, based on the assays, correlate with the activity against ALR2. Figure 5a shows the dose-response curve of compound **14** with an IC_{50} value of 0.22 μM.

Compounds **17** and **18** contain the large fluorene hydrophobic group. This large hydrophobic group is not well accommodated in the hydrophobic pocket of ALR2 and extends into the solvent, perhaps accounting for the somewhat lower affinity of compounds **17** and **18** (15.67 μM and 6.30 μM) relative to compound **14** (0.22 μM). (Figure S6 of Supporting Information and Figure 4b).

On the other hand, compound **22** has a pyrazine group that is a smaller hydrophobic group than either the napthyl group in **14 (**or the fluorene moiety found in compounds **17** and **18)**. The pyrazine group is well accommodated in the ALR2 hydrophobic pocket. However, the pyrazine is too small to make hydrophobic contacts with Trp79, Trp111, Phe122, Phe115 and Leu300 as compound **14** does. This may account for the somewhat weaker activity compound **22**(3.65 μM, Figure 5c) than **14 (**0.22 μM**)**.

Compound **7** demonstrated moderate activity against ALR2 with an IC_{50} value of 25.05 μM (Table 1). The predicted binding mode of compound **7** is shown in Figure 4d. In this binding mode, compound **7** forms hydrogen bonds with His110 and Trp111 via its ketone oxygen, and another hydrogen bond between its benzofuran-5-hydroxyl and Trp20. In addition, π-π stacking interactions between the benzene ring of compound **7** and Trp111 were observed, which may well contribute to the activity.[23] One of the reasons for the π-π stacking interactions is the orientation of the three

hydrogen bonds. MD simulations indicated the RMSD of compound **7** fluctuates around 1 Å (Figure S5 of Supporting Information).

It has been reported that sulfonic or sulfonamide groups on an ALR2 inhibitor form hydrogen bonds with Tyr48, His110, or Trp111 in the anion binding pocket.[23, 35] This moiety was found in active compounds from the TCM database as well. Compounds **11** and **12,** containing this functional group displayed inhibition of ALR2, with compound **11**, having an IC_{50} of 11.14 μM. Docking studies demonstrated that the binding modes of compounds **11** and **12** were similar and indeed reflected the hydrogen bonds (Figure S6b and 6c of Supporting Information) reported in the previous studies.[23, 35] Interestingly, compound **12** has both carboxyl and sulfonamide groups, and the sulfonamide is found to bind in the anion pocket in the docked structure.

Compound **25** is a curcumin derivative, a natural product extracted from *Curcuma longa* L. Curcumin has been shown to be effective in delaying streptozotocin (STZ)-induced diabetic cataracts in rats mainly through its antioxidant properties and inhibition of ALR2.[47] It inhibited human recombinant ALR2 with an IC_{50} of 10.0 μM.[48] Compound **25** demonstrated 11-fold increased inhibition against ALR2 over that of curcumin itself. The dose-response curve of compound **25,** with an IC_{50} value of 0.89 μM, is depicted in Figure 5c. It was predicted that the 3, 5-dione of compound **25** binds in the anion binding pocket of ALR2, and forms three hydrogen bonds with Tyr48, His110, and Trp111 (Figure 4e). An additional hydrogen bond was formed between the ether and carboxyl groups of compound **25** and Val 299. Also, π-π

stacking interactions between one of the phenyl rings in compound **25** and Trp111 is observed. Like compound **7,** the binding mode of compound 2**5** to ALR2 was similar to that of the ligand IDD594 in the 1US0 structure.

Both the flexibility of ALR2 and the ability of the ligands to induce conformational deformations to accommodate binding are reflected in the binding modes of these new ALR2 inhibitors. As shown in Figure S9 of Supporting Information, ligand induced-fit results in side-chain and even backbone changes in Cys298, Ala299, Leu300, and Phe122.

Selectivity Studies Against ALR1. Many ALR2 inhibitors were potent in both in vitro and in vivo studies, even in animal studies, but still failed in clinical trials due to side effects or poor efficacy. The side effects may be due in part to the failures of selective inhibition of ALR2 with respect to ALR1 (aldehyde reductase, EC 1.1.1.2)[49]. The sequence similarity of ALR1 and ALR2 is close to 65%.[50] To assess the selectivity of the 26 confirmed ALR2 inhibitors given in Figure 3, the 15 most potent compounds were tested for their ability to inhibit human recombinant ALR1 as well (Table 2). The assays confirmed that 8 of the compounds showed a maximum inhibition of ALR1 of less than ~ 25% at 50 μM, with five of these less than 10% (Table 2). Compound **14**, the most active compound, only exhibits inhibition of ALR1 by 24% at 50 μM, which is somewhat better than epalrestat's selectivity (90% inhibition of ALR1 at 50 μM).

IC_{50}'s were determined for the 7 hits which inhibited ALR2 to greater than 40% at 50 μM. The only compound among these ALR2 inhibitors which didn't show

selectivity was compound **7**, IC_{50} values are 31.33 μM for ALR1 (Figure S8) and 25.05 μM for ALR2. As revealed by the IC_{50}, compound **25,** the second most active compound (IC_{50}=0.89 μM), exhibited 100-fold selectivity for ALR2 with respect to ALR1 (IC_{50}=94.65 μM, Table 2). The similar potency and superior selectivity of these compounds when compared to the drug epalresta suggest that these new ALR2 inhibitors presented here are worthy of further development.

Structure Activity Relationship (SAR) for Compound 14 and Analogues. Compound **14**, a β-amino-phenylpropanoic derivative, demonstrates excellent potency and selectivity for ALR2. Actually, compounds **14**, **22**, **23**, and **24** share the same β-amino-phenylpropanoic scaffold with consistent efficacy against ALR2. The latter three compounds show even greater selectivity for ALR2 than **14,** based on the percent inhibition at 50 μM data.

To further explore the SAR of β-amino-phenylpropanoic derivatives for ALR2 inhibition, 10 additional analogues were selected by a substructure search of the GSMTL database in house. The ALR2 inhibition activities were evaluated as listed in Table 3.

The β-amino-phenylpropanoic scaffold has three substituent positions, $R_{1\text{-}3}$. Substituents at R_1 include: naphthalene (**14**); 1,2,3,4-tetrahydronaphthalene (**15**, **29**), pyrazine (**22**-**24**); pyridine (**35**, **36**); and phenyl groups. The activities of these groups are in the following order:

Naphthalene > pyrazine > pyridine > phenyl >1,2,3,4-tetrahydronaphthalene.

This indicates that a large aromatic hydrophobic group at R_1 is important for

inhibition, and that the size and orientation of R_1 plays a key role. Docking studies showed that R_1 resides in the hydrophobic specificity pocket consisting of Trp79, Trp111, Phe122, Phe115 and Leu300. Smaller hydrophobic groups for R_1 show lower activity due to inadequate hydrophobic interactions in the binding pocket. Compound **15** has a large hydrophobic group for R_1, however, the aliphatic portion of the tetrahydronaphthalene group cannot be accommodated as well in the pocket, which also results in the loss of a hydrogen bond between the acyl oxygen of compound **15** with Trp111of ALR2 accounting for its reduced affinity (Figure 6a and 6b). Similar considerations apply to compound **29** (Table 3).

The R_2 group on the benzene ring in the β-amino-phenylpropanoic derivatives is surrounded by hydrophobic residues, Trp20, Val47 and Tyr48 (Figure 6a). It was expected that an additional hydrophobic group at R_2 would further favor activity. This was investigated by assessing the activities of compounds **27**, **28**, **30** and **33** (Table 3)**.** These compounds share the same scaffold, and the only variation is at R_2 (**27**, F; **28**, Cl; **30**, CH_3; **33**, H). The trend here is not so clear. Compounds **27**, **28**, and **30** are more active than compound **33,** which has hydrogen at R_2**,** consistent with the hypothesis, however compounds **27** and **28** are essentially equally active contrary to what might be expected. The SAR of R_3 is unclear due to insufficient data. The general SAR for β-amino-phenylpropanoic scaffold is summarized in Figure 6c.

In Vitro Cytotoxicity Assay. A necessity of any drug compound is that it has negligible toxicity. As a preliminary assessment of toxicity we carried out cell viability assays for the 15 hits which inhibited ALR2 to greater than 50% at 50 μM.

The viability of human embryonic kidney cell lines 293 (HEK 293) with these compounds was evaluated. The results are given in Table 4. With the exception of compounds **11** and **13** the remaining compounds showed negligible cytotoxicity, similar to that of epalrestat. The most active compounds **14** and **2**5 exhibited only 10.46% and 13.85% inhibition at 12.5 μM, respectively. The similar potency, lack of toxicity, and superior selectivity of these compounds when compared to the drug Epalresta suggest that the new ALR2 inhibitors presented here are worthy of further development.

The Pharmacokinetic Properties of the Active Compounds. ADME/T properties, including the absorption, solubility, BBB, hepatotoxicitye, CYP 2D6, alogP, and PSA of the 26 ALR2 inhibitors, have been evaluated *in silico* through Discovery Studio 2.1 (Accelrys, Inc., San Diego, CA). Compared with Epalrestat (the only marketed ALR2 inhibitor), the ADME/T properties of most of our compounds are in the required druggability ranges, especially for compound **14** (β-amino-phenylpropanoic scaffold). The detailed results and comparisons can be found in Table S5 (Supporting Information).

CONCLUSIONS

New selective human aldose reductase 2 (ALR2) inhibitors have been discovered by a protocol of virtual screening in multiple binding pocket conformations followed by binding assays as well as selectivity and toxicity assessments. Several of the resulting potent ALR2 inhibitors contain a β-amino-phenylpropanoic scaffold. Biological tests demonstrated that two of the ALR2 inhibitors demonstrate

sub-micromolar IC_{50} values (0.22 μM and 0.89 μM), which are comparable to the commercially available drug, epalrestat, while showing superior selectivity to epalrestat relative to inhibition of ALR1. SAR studies on the β-amino-phenylpropanoic compound **14** and its analogs provide insight for further optimization of these leads.

These results also tend to validate the *in silico* methods used. While conventional structure-based virtual screening often starts with single conformation, we have exploited a conformational sampling process via MD simulations. The conformation sampling process produces multiple consensus binding pockets, which represent the dynamic conformation changes of a binding pocket interacting with a different ligand. A series of three filters were then used on the hits to optimize the yields. Our screening protocol applied to this system resulted in more active hits and a success rate of 14% based on compounds retrieved with IC_{50}'s of 10 μM or less. Recently, many tools and protocols have been developed in our labs for lead identification;[51-54] we will combine these tools and protocols to improve the quality of ALR2 inhibitors.

EXPERIMENTAL SECTION

Receptor Preparation. Three ALR2 protein-ligand complexes (PDB entries: 2PDK, 1US0, and 2FZD) were placed in a TIP3P octahedral water box such that the box boundary is at least 10 Å from any protein atom, and MD simulations were performed for 6 nanoseconds. The water molecules in the crystal data were kept. Three average structures were calculated from the equilibrated stage of the MD trajectories (from 5-6 nanoseconds) of 2PDK, 1US0 and 2FZD. The three averaged

structures were then optimized with the steepest descent method for 200 steps using Discovery Studio 2.1.[55]

The FlexX module in Sybyl 7.3[56] was used to dock our compound library against the three consensus structures. The active sites were defined as all residues within 6.5 Å radius of the reference molecule in each optimized structure and cofactors were retained during the docking process. Other FlexX parameters were set to default values.

Compound Library Preparation. The structures from the GSMTL database (~7,249 compounds) were docked against the three consensus structures. All inorganic atoms in the structures were removed prior to the virtual screening using the MOE package (Molecular Operating Environment).[39] The remaining moieties were preprocessed (eg, adding hydrogen atoms; setting ionization states to be appropriate for a pH range of 6.5 to 8.5; and generating stereo isomers and valid single 3D conformers) by means of the Ligand Preparation module of Discovery Studio 2.1.

The structures of the training compounds and their inhibition data (IC_{50}< 50 μM) were downloaded from the Binding DB database,[38] and were preprocessed as described above.

Establishment of the Selection Criteria. 927 compounds, which had IC_{50}'s lower than 50 μM, were selected from the Binding DB database. The compounds were individually docked against the three consensus structures with FlexX. This results in three FlexX scores for each compound: Fscore1 (derived from 2PDK), Fscore2 (derived from 1US0), and Fscrore3 (derived from 2FZD). Thus three sets of 927

FlexX scores were obtained for the 927 compounds docked to each of the three protein structures. One of the selection criteria was obtained by averaging each set of docking scores for a given structure, i.e. $FS_i = \langle Fscore_i \rangle_{927}$; i=1-3. Then criterion 1 was invoked.

Criterion 1. Each of the hit compound's three FlexX scores (generated from docking into 2PDK, 1US0, and 2FZD respectively) should be more favorable than the corresponding average scores obtained for training set for each system.

The second selection criterion was obtained by insisting that certain key protein-ligand interactions were satisfied in the docked structure. The key ligand binding residues were identified by means of MOE. Protein ligand interaction fingerprints (PLIF) were calculated from 76 ALR2 structures selected from the PDB. This corresponded to all ligand ALR2 structures excluding those containing mutations in the active site. From this analysis it was concluded that:

Criterion 2. A hit compound's docked conformation should have hydrogen bonds to at least two of the three key residues (i.e., Tyr48, His110, and Trp111) and hydrophobic interaction with the specificity hydrophobic pocket (Trp79, Trp111, Phe122, Phe115 and Leu300).

Finally the third criterion simply insures that the dynamic trajectory hasn't taken the ligand-protein to another conformational state or pose of the ligand.

Criterion 3: The RMSD of the ligand, from its docked pose, resulting from the MD simulation should be less than 3 Å.

A final virtual screening hit must satisfy all three criteria.

Virtual Screening. The FlexX docking program (Sybyl 7.3) was employed for the virtual screening. As a consistency check three native ligands were docked back into their host crystal structures to confirm the performance of FlexX for the human ALR2 system. The RMSD between the docking pose of native ligand and the experimental pose in the consensus structure complex was less than 1 Å.

The 927 training compounds derived from the Binding DB database were docked to the three consensus structures, respectively. Only the best pose per ligand, as described above, was recorded for each docking run.

Structures from the GSMTL database were then virtually screened against the three consensus structures simultaneously with the FlexX docking program. For each compound, the program recorded 20 structural poses per run. These poses were analyzed by a clustering algorithm, and the structure from the largest cluster with the best FlexX score was selected. If the docking score was more favorable than the average docking score of known hits to that structure, ie $FS_i \leq \langle Fscore_i \rangle_{927}$) as defined above, the corresponding compound was then assessed by the two final criteria.

MD Simulations. MD simulations were performed using the PMEMD module in AMBER 11[57] accelerated by running on a GPU system, the NVIDIA CUDA processor. The three initial co-crystal structures (2PDK, 1US0 and 2FZD) were solvated in a 10Å octahedral box with TIP3P water. Periodic boundary conditions were applied.

The AMBER ff99SB force field[58] was exploited for the protein, and the general AMBER force field (GAFF)[59] was applied to the three structures. The cofactor

parameters were obtained from the literature.[60] The partial charges of three substrates were computed using the HF/6-31 G* basis set from GAUSSIAN03,[61] and refined by RESP calculation using the antechamber module of the AMBER 11 package. Sodium ions were added in order to neutralize the systems. To remove possible steric stresses, the systems were minimized for 1,000 steps with the steepest descent method, followed by application of conjugate gradients for another 1,000 steps. Each of the three systems was linearly heated from 0 to 300K using a Langevin thermostat, with a collision frequency of 5.0 ps^{-1}, and harmonic restraints of 4 kcal/mol/$Å^2$ on the backbone atoms over 20 ps and then equilibrated for 50 ps at 300 K using the NVT ensemble. Finally, dynamics simulations of 6 ns were carried out for the production step in an NPT ensemble at 1atm and 300 K. The coordinates of the system were saved at every picosecond.

The temperature was kept at 300 K by means of a weak coupling algorithm.[62] Covalent bonds involving hydrogen were constrained using the SHAKE algorithm.[63] The Particle-Mesh-Ewald method[64] was applied to treat the long range electrostatic interactions with a 10 Å non-bonded cutoff. The three average structures were calculated from the equilibrated stage of the MD trajectories (from 5 ns to 6 ns), and subsequent optimized with steepest descents for 200 steps. The three minimized average structures were then used for the virtual screening campaign. 128 structural complexes, which satisfied the first two filters, were submitted for MD simulations as described. The RMSD between the docked pose and averaged MD simulated pose for a hit candidate was checked to make sure the ligand pose had not undergone a

conformational transition in the MD (criterion 3) to reduce the chances of false positives from the docking procedure.

Chemistry. All compounds tested for ALR2 inhibitions were taken from the GSMTL in our laboratory. Purity of the compounds was assessed by HPLC equipped with a ZORBAX SB-C18 column (250 mm×4.6 mm, 5 μm particle size) and a UV/VIS detector setting of λ = 254 nm. The compounds were eluted with the two solvent systems (CH_3OH as the organic phase in method I and CH_3CN as the organic phase in method II) at a flow rate of 0.5 mL/min. HPLC analysis of the compounds assayed confirmed the purity at ≥ 95% (Table S3). Sources information and ^1H NMR, MS data were listed in Table S4.

Expression and Purification of Recombinant Human ALR2. Recombinant human ALR2 was expressed and purified as described by Nishimura et al.[65] The human ALR2 gene was cloned into pET15b vector (Novagen) and expressed in *Escherichia coli* strain BL21 (+) (Novagen). The hexahistidine tagged protein was induced by IPTG (Isopropyl β-D-1-thiogalactopyranoside) during a 20 h period at 25 °C and purified using a Ni-affinity column (Qiagen).

Expression and Purification of Recombinant Human ALR1. Expression and purification of recombinant human ALR1 was carried out following the methods described by Bohren *et al*.[66, 67] The recombinant human ALR1 expression plasmid in pReceiver-B01 was expressed in *E. coli* using a T7-based expression system and purified using a Ni-affinity column.

In Vitro Recombinant Human ALR2 Inhibition Assay. A spectrophotometric

assay[68] was employed for in vitro inhibitory tests via the detection of absorbance decreases from the oxidation of NADPH to $NADP^+$ catalyzed by the ALR2 enzyme. The absorbance at 340 nm was monitored at 30℃ with an ultraviolet spectrophotometer reader. The assay was performed using a 1 mL reaction curette with sodium phosphate buffer (0.1 M), NADPH cofactor (0.15 mM), Li_2SO_4 (0.4 M), human ALR2 (0.486 μM), and D-glyceraldehyde substrate (10 mM).

The predicted hits and reference compounds (quercetin and epalrestat) were dissolved in dimethyl sulfoxide (DMSO). The final concentration of DMSO was not more than 1%. The inhibitory activities of tested compounds were assayed by adding them to the reaction cuvettes at 50 μM. Those compounds found to be active were tested at additional concentration ranging from 0.1 to 10 μM. The IC_{50} value for each compound was determined as the compound concentration that inhibited human ALR2 activity by 50%. The IC_{50} values was curve-fitted as described by Alexiou et al.[69] Compounds were tested at a minimum of five concentrations and all experiments were performed in triplicate. To exclude any possible nonspecific/promiscuous inhibition of ALR2, we deepened our hit validation, repeating all the assays in the presence of 0.01% Triton X-100, as suggested by Shoichet.[63] None of the observed inhibitory activities was affected by the addition of the nonionic detergent, confirming the activity.

In Vitro Recombinant Human ALR1 Inhibition Assay. To study the selectivity of ALR2 inhibitors, an inhibition study of ALR1 was carried out by monitoring the oxidation of NADPH at 340 nm as a function of time using glyceraldehyde as

substrate. The assay mixture contained 0.1 M sodium phosphate buffer of pH 6.2, 10 mM DL-glyceraldehyde, 0.1 mM NADPH. The confirmed ALR2 inhibitors were added to the ALR1 assay mixture and tested at 37℃. The IC_{50} values of compounds having > 40% inhibition at 50 μM were determined as previously described.[69] Compounds were tested at a minimum of five concentrations and the experiments were performed in triplicate.

In Vitro Cell Viability Assay. Cell proliferation was measured in an MTT assay protocol (Table 4). Five thousand HEK 293 (human embryonic kidney 293) cells were inserted in a 96-well plate in 100 μL of indicated medium in the presence of compounds at the indicated concentration. After incubation for 48 hours, the cells were further incubated with 20 μL of 2.5 mg/mL MTT for 4 hours at 37 ℃ in a humidified incubator with 5% CO_2. Then the formazan dye was dissolved in 100 μL of DMSO, and the absorbance was measured at 570 nm by using PowerWave^{TM}XS microplate spectrophotometer (BioTek). The inhibition rate (%) was calculated as:

$$\text{Inhibition (\%)}=[1-(A_{570,\,compd})/(A_{570,\,control})]\times 100\%$$

ASSOCIATED CONTENT

Supporting Information: HPLC, MS, and ^{1}H NMR data, MD simulation results, docking results, and dose response results, and additional predicted binding modes of hits 11, 12 and 17 are resented in the supplementary material. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author: (QG) Phone: +86-20-39943077. Fax: +86-20-39943077. Email: guqiong@mail.sysu.edu.cn. (JX) Phone: +86-20-39943023. Fax: +86-20-39943023. Email: junxu@biochemomes.com. (AH): Email: athagler@gmail.com

Notes: The authors declare no competing financial interest.

ACKNOWLEDGMENT. We gratefully acknowledge Guangdong Small Molecule Tangible Library (GSMTL) for supplying database and compounds for bioassay. This work was funded in part of the National High-tech R&D Program of China (863 Program) (2012AA020307), the National Science and Technology Major Project of China (2010ZX09102-305), Guangdong Recruitment Program of Creative Research Groups, the National Natural Science Foundation of China (No. 81001372, 81173470).

REFERENCES

(1) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N. Engl. J. Med.* **1993**, *14*, 977-986.

(2) Eisenmann M, Steuber H, Zentgraf M, Altenkamper M, Ortmann R, Perruchon J, Klebe GSchlitzer M. Structure-based optimization of aldose reductase inhibitors originating from virtual screening. *ChemMedChem* **2009**, *5*, 809-819.

(3) Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **2001**, *6865*, 813-820.

(4) Kador P F, Kinoshita J HSharpless N E. Aldose reductase inhibitors: a potential new class of agents for the pharmacological control of certain diabetic complications. *J. Med. Chem.* **1985**, *7*, 841-849.

(5) Kinoshita J HNishimura C. The involvement of aldose reductase in diabetic complications. *Diabetes Metab. Rev.* **1988**, *4*, 323-337.

(6) Bhatnagar ASrivastava S K. Aldose reductase: congenial and injurious profiles of an enigmatic enzyme. *Biochem. Med. Metab. Biol.* **1992**, *2*, 91-121.

(7) Alexiou P, Pegklidou K, Chatzopoulou M, Nicolaou IDemopoulos V J. Aldose reductase enzyme and its implication to major health problems of the 21(st) century. *Curr. Med. Chem.* **2009**, *6*, 734-752.

(8) Ramana K VSrivastava S K. Aldose reductase: a novel therapeutic target for inflammatory

pathologies. *Int. J Biochem. Cell. Biol.* **2010**, *1*, 17-20.
(9) Ramana K V. ALDOSE REDUCTASE: New Insights for an Old Enzyme. *Biomol. Concepts* **2011**, *1-2*, 103-114.
(10) Iwata Y, Arisawa M, Hamada R, Kita Y, Mizutani M Y, Tomioka N, Itai AMiyamoto S. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J. Med. Chem.* **2001**, *11*, 1718-1728.
(11) Dvornik E, Simard-Duquesne N, Krami M, Sestanj K, Gabbay K H, Kinoshita J H, Varma S DMerola L O. Polyol accumulation in galactosemic and diabetic rats: control by an aldose reductase inhibitor. *Science* **1973**, *117*, 1146-1148.
(12) Sestanj K, Bellini F, Fung S, Abraham N, Treasurywala A, Humber L, Simard-Duquesne NDvornik D. N-[5-(trifluoromethyl)-6-methoxy-1-naphthalenyl]thioxomethyl]- N-methylglycine (Tolrestat), a potent, orally active aldose reductase inhibitor. *J. Med. Chem.* **1984**, *3*, 255-256.
(13) Kikkawa R, Hatanaka I, Yasuda H, Kobayashi N, Shigeta Y, Terashima H, Morimura TTsuboshima M. Effect of a new aldose reductase inhibitor, (E)-3-carboxymethyl-5-[(2E)-methyl-3-phenylpropenylidene]rhodanine (ONO-2235) on peripheral nerve disorders in streptozotocin-diabetic rats. *Diabetologia* **1983**, *4*, 290-292.
(14) Mylari B L, Larson E R, Beyer T A, Zembrowski W J, Aldinger C E, Dee M F, Siegel T WSingleton D H. Novel, potent aldose reductase inhibitors: 3,4-dihydro-4-oxo-3-[[5-(trifluoromethyl)-2-benzothiazolyl] methyl]-1-phthalazineacetic acid (zopolrestat) and congeners. *J. Med. Chem.* **1991**, *1*, 108-122.
(15) Ao S, Shingu Y, Kikuchi C, Takano Y, Nomura K, Fujiwara T, Ohkubo Y, Notsu YYamaguchi I. Characterization of a novel aldose reductase inhibitor, FR74366, and its effects on diabetic cataract and neuropathy in the rat. *Metabolism* **1991**, *1*, 77-87.
(16) Stribling D, Mirrlees D J, Harrison H EEarl D C. Properties of ICI 128,436, a novel aldose reductase inhibitor, and its effects on diabetic complications in the rat. *Metabolism* **1985**, *4*, 336-344.
(17) Van Zandt M C, Jones M L, Gunn D E, Geraci L S, Jones J H, Sawicki D R, Sredy J, Jacot J L, Dicioccio A T, Petrova T, Mitschler APodjarny A D. Discovery of 3-[(4,5,7-trifluorobenzothiazol-2-yl)methyl]indole-N-acetic acid (lidorestat) and congeners as highly potent and selective inhibitors of aldose reductase for treatment of chronic diabetic complications. *J. Med. Chem.* **2005**, *9*, 3141-3152.
(18) Schemmel K E, Padiyara R SD'Souza J J. Aldose reductase inhibitors in the treatment of diabetic peripheral neuropathy: a review. *J. Diabetes. Complications.* **2010**, *5*, 354-360.
(19) Sotriffer C A, Kramer OKlebe G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* **2004**, *1*, 52-66.
(20) Steuber H, Heine A, Podjarny AKlebe G. Merging the binding sites of aldose and aldehyde reductase for detection of inhibitor selectivity-determining features. *J. Mol. Biol.* **2008**, *5*, 991-1016.
(21) Howard E I, Sanishvili R, Cachau R E, Mitschler A, Chevrier B, Barth P, Lamour V, Van Zandt M, Sibley E, Bon C, Moras D, Schneider T R, Joachimiak APodjarny A. Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 A. *Proteins* **2004**, *4*, 792-804.
(22) Steuber H, Zentgraf M, Gerlach C, Sotriffer C A, Heine AKlebe G. Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. *J. Mol. Biol.* **2006**, *1*, 174-187.

(23) Cosconati S, Marinelli L, La Motta C, Sartini S, Da Settimo F, Olson A JNovellino E. Pursuing Aldose Reductase Inhibitors through in Situ Cross-Docking and Similarity-Based Virtual Screening. *J. Med. Chem.* **2009**, *18*, 5578-5581.

(24) Alonso H, Bliznyuk A AGready J E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *5*, 531-568.

(25) Liu W G, Schmidt B, Voss GMuller-Wittig W. Molecular dynamics simulations on commodity GPUs with CUDA. *High Performance Computing - Hipc 2007, Proceedings* **2007**, 185-196
663.

(26) Grand S L G, Andreas W; Xu, Dong; Poole, Duncan; Walker, Ross C. . Acceleration of amber generalized born calculations using nvidia graphics processing units. *http://ambermd.org/gpus/* **2010**.

(27) Osguthorpe D J, Sherman WHagler A T. Generation of Receptor Structural Ensembles for Virtual Screening Using Binding Site Shape Analysis and Clustering. *Chem. Biol. Drug. Des.* **2012**. *80*, 182-193.

(28) Osguthorpe D J, Sherman, W., and Hagler, A. T. Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols. *J. Phys. Chem. B.* **2012**. *116,* 6952-6959.

(29) Klebe G, Kramer OSotriffer C. Strategies for the design of inhibitors of aldose reductase, an enzyme showing pronounced induced-fit adaptations. *Cell. Mol. Life. Sci.* **2004**, *7-8*, 783-793.

(30) Klebe G, Zeragraf M, Steuber H, Koch C, La Motta C, Sartini SSotriffer C A. How reliable are current docking approaches for structure-based drug design? Lessons from aldose reductase. *Angew. Chem.-Int. Edit.* **2007**, *19*, 3575-3578.

(31) Lyne P D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *20*, 1047-1055.

(32) De Azevedo W F, Jr. Structure-based virtual screening. *Curr. Drug Targets* **2010**, *3*, 261-263.

(33) Villoutreix B O, Eudes RMiteva M A. Structure-based virtual ligand screening: recent success stories. *Comb. Chem. High Throughput Screening* **2009**, *10*, 1000-1016.

(34) Kraemer O, Hazemann I, Podjarny A DKlebe G. Virtual screening for inhibitors of human aldose reductase. *Proteins* **2004**, *4*, 814-823.

(35) Rastelli G, Ferrari A M, Costantino LGamberini M C. Discovery of new inhibitors of aldose reductase from molecular docking and database screening. *Bioorg. Med. Chem.* **2002**, *5*, 1437-1450.

(36) Sotriffer C ADramburg I. "In situ cross-docking" to simultaneously address multiple targets. *J. Med. Chem.* **2005**, *9*, 3122-3125.

(37) Gu Q, Xu J, Gu L. Selecting diversified compounds to build a tangible library for biological and biochemical assays. *Molecules* **2010**, *7*, 5031-5044.

(38) Liu T, Lin Y, Wen X, Jorissen R NGilson M K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, 198-201.

(39) Audie JScarlata S. A novel empirical free energy function that explains and predicts protein-protein binding affinities. *Biophys Chem.* **2007**, *2-3*, 198-211.

(40) Wirasathien L, Pengsuparp T, Suttisri R, Ueda H, Moriyasu MKawanishi K. Inhibitors of aldose reductase and advanced glycation end-products formation from the leaves of Stelechocarpus cauliflorus R.E. Fr. *Phytomedicine* **2007**, *78*, 546-550.

(41) Jung H A, Yoon N Y, Kang S S, Kim Y SChoi J S. Inhibitory activities of prenylated flavonoids from Sophora flavescens against aldose reductase and generation of advanced glycation endproducts. *J. Pharm. Pharmacol.* **2008**, *9*, 1227-1236.

(42) Xu J. ^{13}C NMR Spectral Prediction by Means of Generalized Atom Center Fragment Method. *Molecules* **1997**, *8*, 114-128.
(43) Yan X, Gu Q, Lu F, Li JXu J. GSA: a GPU-accelerated structure similarity algorithm and its application in progressive virtual screening. *Mol. Diversity* **2012**, *4*, 759-769.
(44) Nidhi Sharma R K G, Anuradha Kaplish, Chander Mohan. Flavonoids As Promising Anti-Cancer Agent: A Review. *Int. J. Food. Sci. Tech.* **2012**, 1-154.
(45) Fisanick W. The chemical abstracts service generic chemical (Markush) structure storage and retrieval capability 0.1 Basic concepts. *J. Chem. Inf. Comput. Sci.* **1990**, 145-154.
(46) Kaur G A, K.; Chander, M. Flavonoids as promising anti-cancer agent: a review. *Int. J. Natural Product Sci.* **2012**. *1*, 154.
(47) Suryanarayana P, Saraswat M, Mrudula T, Krishna T P, Krishnaswamy KReddy G B. Curcumin and turmeric delay streptozotocin-induced diabetic cataract in rats. *Invest. Ophthalmol. Vis. Sci.* **2005**, *6*, 2092-2099.
(48) Muthenna P, Suryanarayana P, Gunda S K, Petrash J MReddy G B. Inhibition of aldose reductase by dietary antioxidant curcumin: mechanism of inhibition, specificity and significance. *FEBS Lett.* **2009**, *22*, 3637-3642.
(49) El-Kabbani O, Carbone V, Darmanin C, Oka M, Mitschler A, Podjarny A, Schulze-Briese CChung R P. Structure of aldehyde reductase holoenzyme in complex with the potent aldose reductase inhibitor fidarestat: implications for inhibitor binding and selectivity. *J. Med. Chem.* **2005**, *17*, 5536-5542.
(50) El-Kabbani O, Wilson D K, Petrash MQuiocho F A. Structural features of the aldose reductase and aldehyde reductase inhibitor-binding sites. *Mol. Diversity* **1998**, *4*, 19.
(51) Huang D, Gu Q, Ge H, Ye J, Salam N K, Hagler A, Chen H, Xu J. On the value of homology models for virtual screening: discovering hCXCR3 antagonists by pharmacophore-based and structure-based approaches. *J. Chem. Inf. Model.* **2012**, *5*, 1356-1366.
(52) Fang J, Huang D, Zhao W, Ge, H, Luo H, and Xu J. A new protocol for predicting novel GSK-3β ATP competitive inhibitors. *J. Chem. Inf. Model.* **2011**, 1431-1438.
(53) Zhao W, Gu Q, Wang L, Ge H, Li J, Xu J. Three-dimensional pharmacophore modeling of liver-X receptor agonists. *J. Chem. Inf. Model.* **2011**, *9*, 2147-2155.
(54) Yan X, L Jiabo, Liu Zhihong, Zheng Minghao, Ge Hu, Xu Jun. Enhancing Molecular Shape Comparison by Weighted Gaussian Functions. *J. Chem. Inf. Model.* **2013.** (Just Accepted).
(55) Prakhov N D, Chernorudskiy A LGainullin M R. VSDocker: a tool for parallel high-throughput virtual screening using AutoDock on Windows-based computer clusters. *Bioinformatics* **2010**, *10*, 1374-1375.
(56) Kramer B, Rarey MLengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *2*, 228-241.
(57) D.A. Case T A D, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. AMBER 11, University of California, San Francisco. **2010**.
(58) Hornak V, Abel R, Okur A, Strockbine B, Roitberg ASimmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *3*, 712-725.
(59) Mukherjee G, Patra N, Barua PJayaram B. A fast empirical GAFF compatible partial atomic

charge assignment scheme for modeling interactions of small molecules with biomolecular targets. *J. Comput. Chem.* **2010**. *32*, 893-907.

(60) Sotriffer C A, Kramer OKlebe G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* **2004**, *1*, 52-66.

(61) M.J. Frisch G W T, H.B. Schlegel, G.E. Scuseria, M.A. Robb. Gaussian 03, Revision E.01, Gaussian, Inc, Pittsburgh PA. **2004**.

(62) H.J.C. Berendsen J P M P, W.F. van Gunsteren, A. DiNola, J.R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 3684-3690.

(63) S. Miyamoto P A K. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, 8952-8962.

(64) T. Darden D Y, L. Pedersen. Particle mesh Ewald: an N · log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 10089-10092.

(65) Nishimura C, Yamaoka T, Mizutani M, Yamashita K, Akera TTanimoto T. Purification and characterization of the recombinant human aldose reductase expressed in baculovirus system. *Biochim. Biophys. Acta.* **1991**, *2*, 171-178.

(66) Barski O A, Gabbay K H, Grimshaw C EBohren K M. Mechanism of human aldehyde reductase: characterization of the active site pocket. *Biochemistry* **1995**, *35*, 11264-11275.

(67) Bohren K M, Page J L, Shankar R, Henry S PGabbay K H. Expression of human aldose and aldehyde reductases. Site-directed mutagenesis of a critical lysine 262. *J. Biol. Chem.* **1991**, *35*, 24031-24037.

(68) Stefek M, Snirc V, Djoubissie P O, Majekova M, Demopoulos V, Rackova L, Bezakova Z, Karasu C, Carbone VEl-Kabbani O. Carboxymethylated pyridoindole antioxidants as aldose reductase inhibitors: Synthesis, activity, partitioning, and molecular modeling. *Bioorg. Med. Chem.* **2008**, *9*, 4908-4920.

(69) Alexiou PDemopoulos V J. A Diverse Series of Substituted Benzenesulfonamides as Aldose Reductase Inhibitors with Antioxidant Activity: Design, Synthesis, and in Vitro Activity. *J. Med. Chem.* **2010**, *21*, 7756-7766.

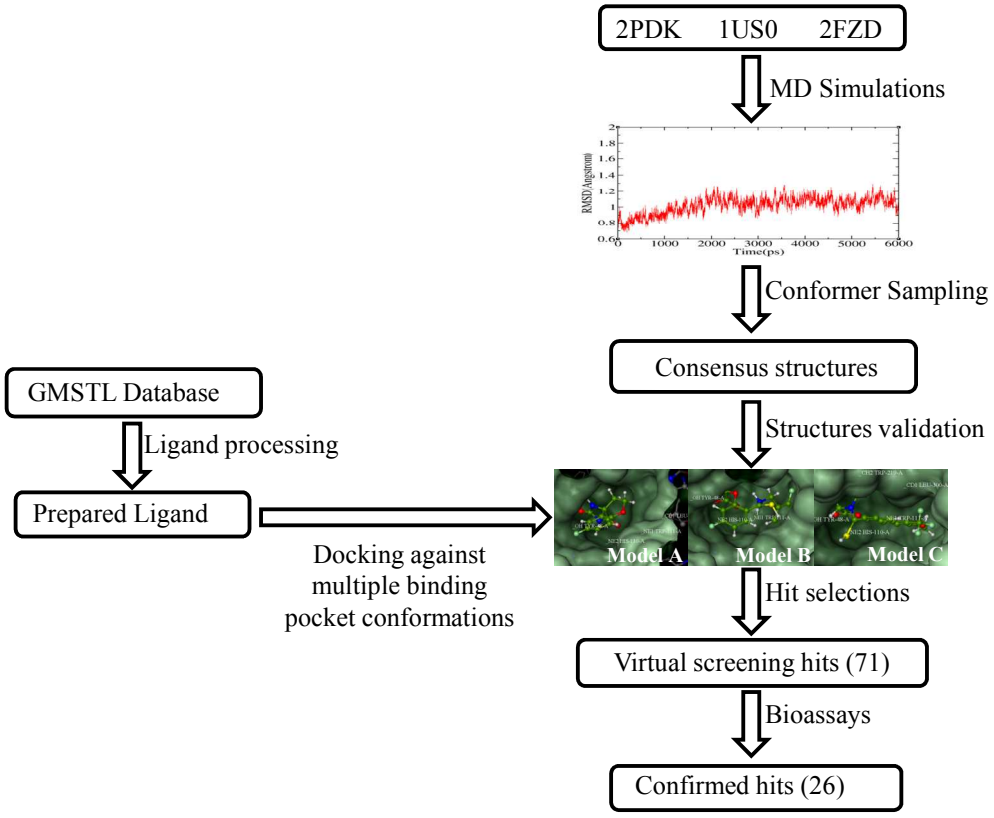


Figure 1. A flowchart of the protocol of virtual screening multiple binding pocket conformations supported by molecular dynamic simulations.

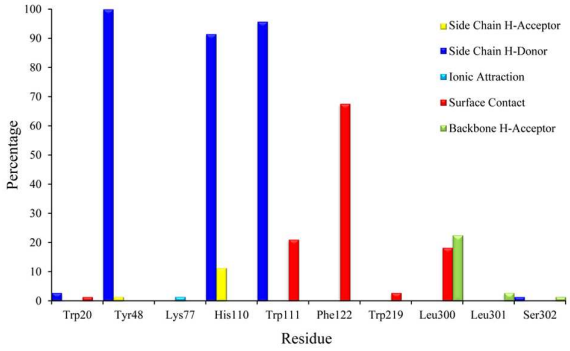


Figure 2. The protein ligand interaction fingerprints (PLIF) derived from 76 available co-crystal structures of ALR2 and inhibitors.

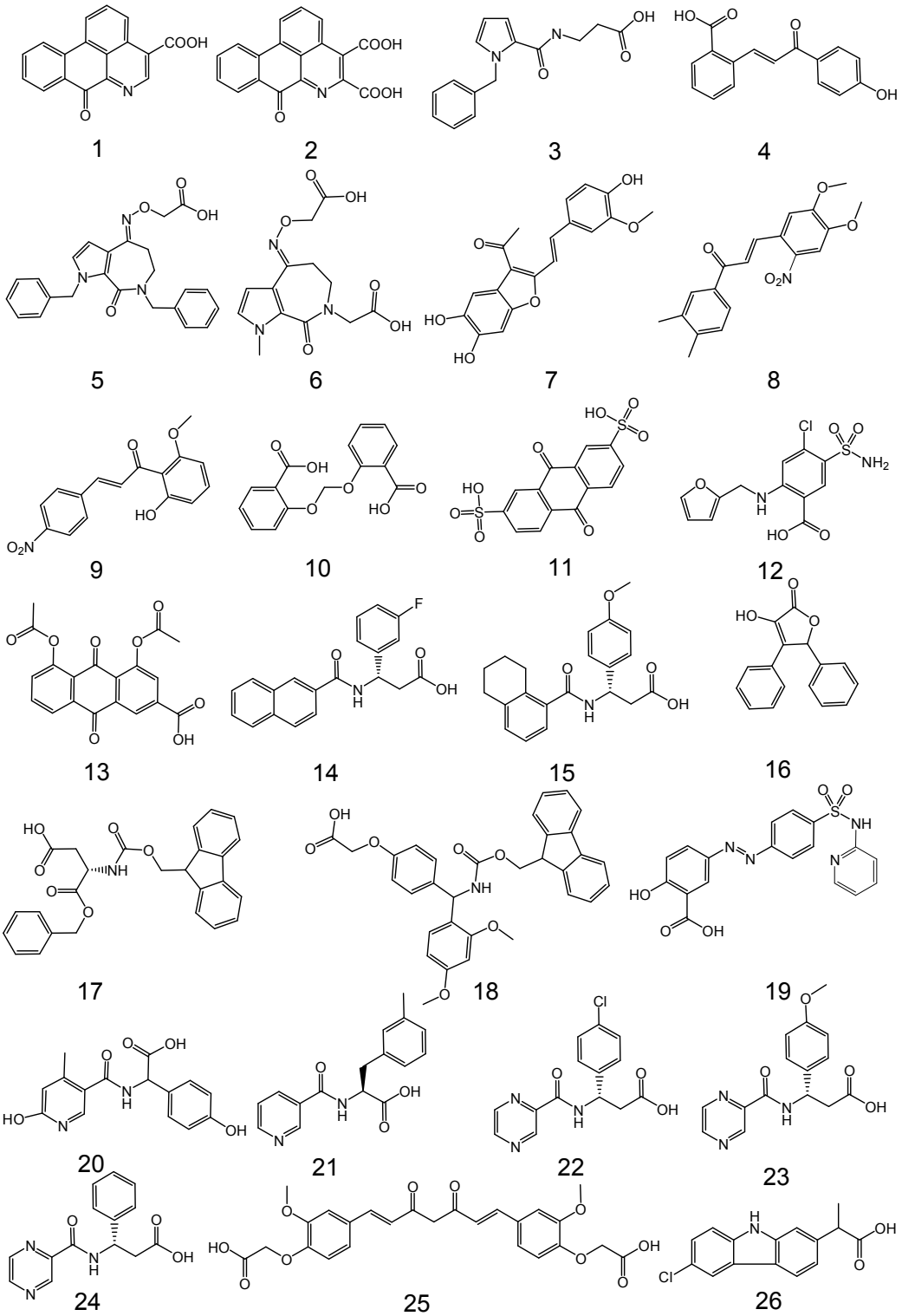


Figure 3. Structures of hits from the virtual screening scheme.

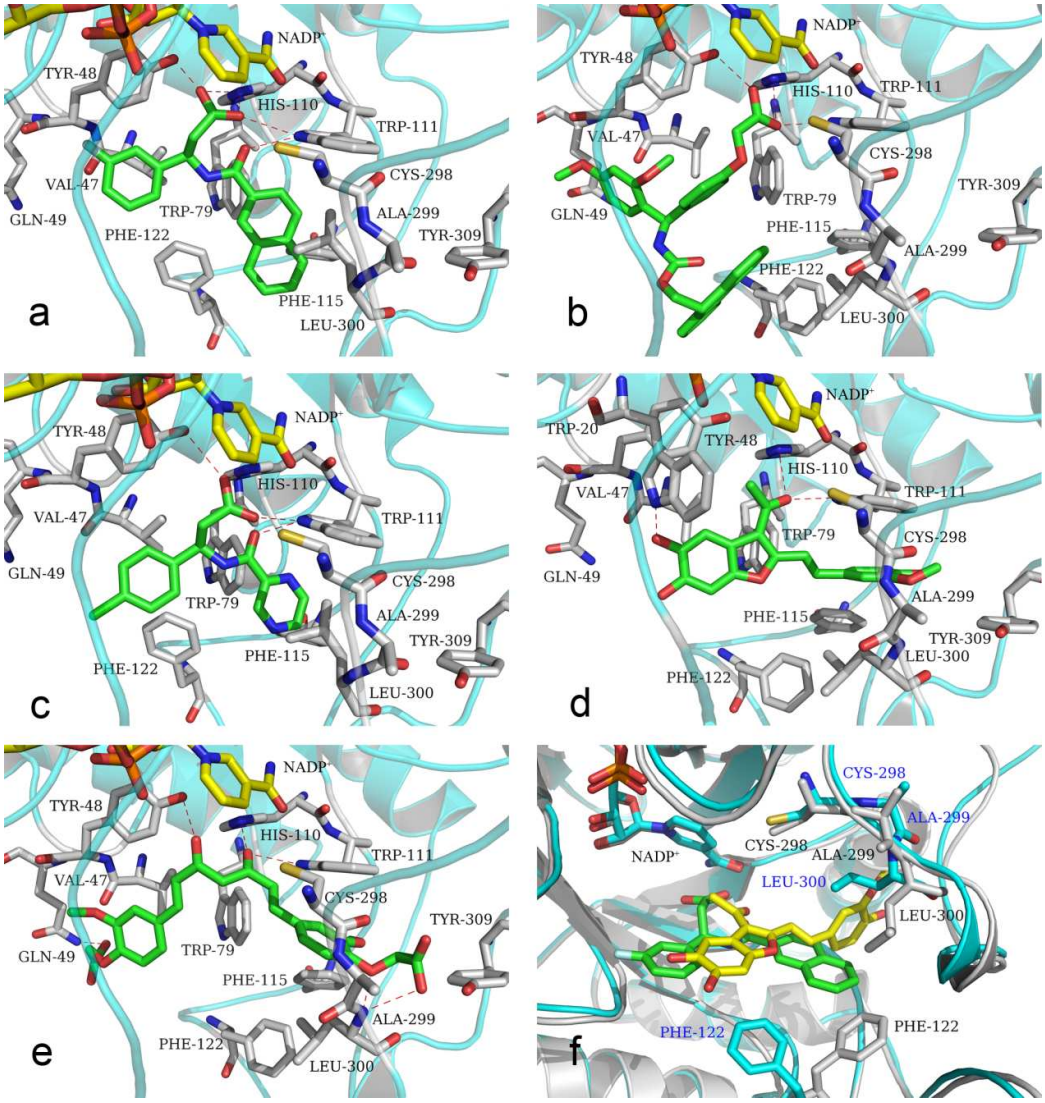


Figure 4. Binding modes for compounds **14** (a), **18** (b), **22** (c), **7** (d), and **25** (e). Hydrogen bonds are depicted by red dotted lines. The induced-fit effect can be seen in (f), where one can see PHE-122, ALA-299 and LEU-300 undergoing significant displacements to accommodate the bound ligand. The bound structures of compounds **7** and **14** are derived by averaging the MD simulated structures from 1US0 and 2FZD, respectively.

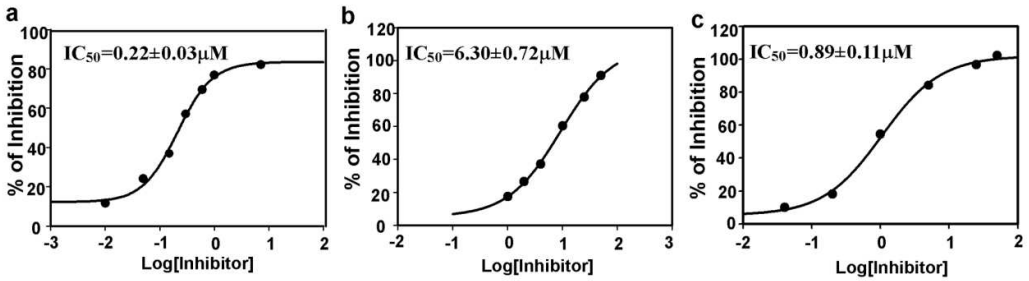


Figure 5. ALR2 inhibition dose-response curves for compound **14** (a), **18** (b), and **25** (c). Values are generated using at least five concentrations of the inhibitors (μM), with triplicate determinations at each concentration. Percent inhibition is plotted on the ordinate, against the log of the concentration on the abcissa.

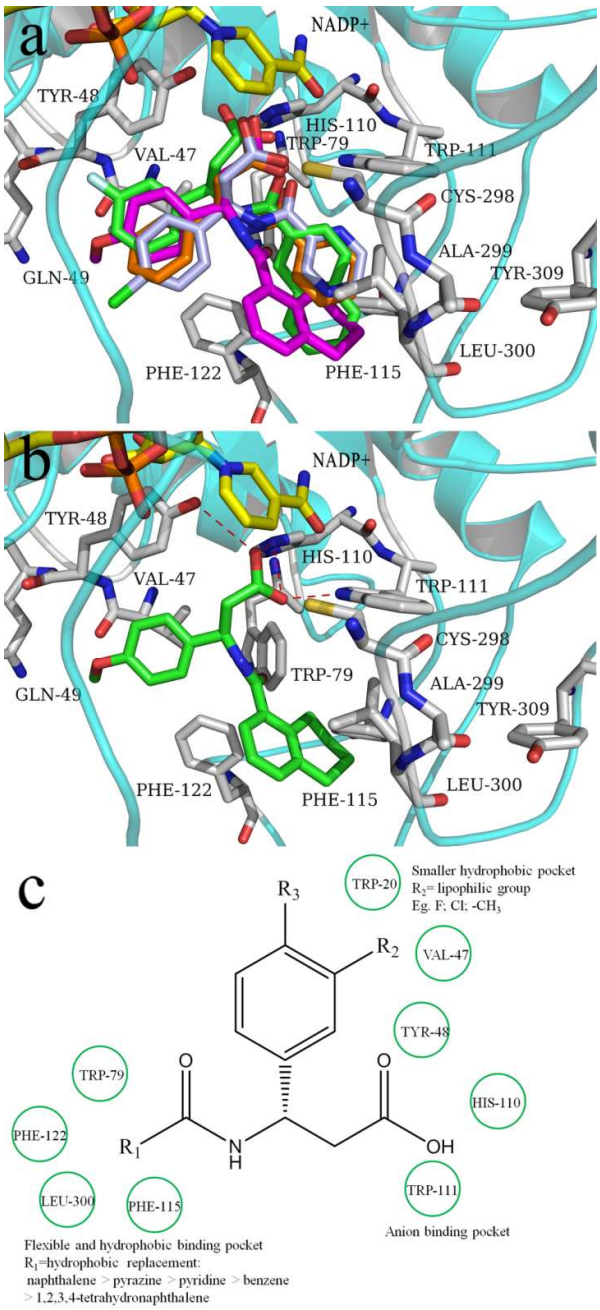


Figure 6. (a) Superimposition of the binding modes of compounds **14** (green), **15** (magenta), **22** (light blue), and **27** (orange). (b) Binding mode for compound **15**. (c) A summary of the SAR of β-amino-phenylpropanoic acid derivates.

Table 1. Virtual screening hits and their in vitro assay results for ALR2 inhibitions.

Compd.	Source	% of ALR2 inhibition at 50μM [a]	IC_{50}(μM) [b]
1	SYSU-00124	76.17±6.2	10.00±0.48
2	SYSU-00123	35.57±3.7	ND
3	SYSU-00295	44.87±2.9	ND
4	SYSU-01227	53.74±7.6	37.79±0.12
5	SYSU-00298	68.80±6.1	31.8±3.38
6	SYSU-00300	40.52±3.1	ND
7	SYSU-00486	88.50±1.8	25.05±3.43
8	SYSU-01606	31.30±5.2	ND
9	SYSU-01809	36.40±3.4	ND
10	SYSU-21694S	80.70±1.9	10.2±1.83
11	SYSU-20957S	68.62±2.2	11.14±0.54
12	SYSU-20665S	46.60±4.1	ND
13	SYSU-10135N	56.76±4.9	26.55±3.29
14	SYSU-22363S	94.91±2.4	0.22±0.03
15	SYSU-22410S	34.36±6.8	ND
16	SYSU-21294S	32.65±3.7	ND
17	SYSU-21741S	72.30±4.3	15.67±2.76
18	SYSU-22315S	81.67±1.2	6.30±0.72
19	SYSU-20433S	83.61±1.6	10.03±2.33
20	SYSU-22133S	30.08±5.3	ND
21	SYSU-22424S	44.65±3.4	ND
22	SYSU-22433S	83.25±2.7	3.65±0.26
23	SYSU-22439S	70.58±4.5	20.20±1.74
24	SYSU-22449S	86.72±4.1	4.3±1.20
25	SYSU-00241	94.30±1.8	0.89±0.11
26	SYSU-20215S	45.60±2.9	ND
Quercetin[c]		74.10±2.2	19.24±1.04
Epalrestat[c]		92.36±4.3	0.24±0.01

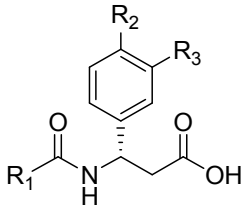
[a] % Inhibition values are the mean ± SD of triplicate measurements at 50μM. [b] IC_{50} values for ALR2 shown are the mean ± SD of triplicate measurements. [c] Used as positive control compounds. ND: not determined.

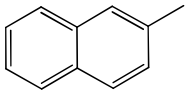
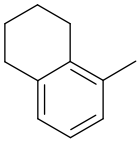
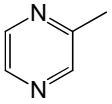
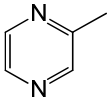
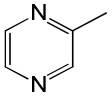
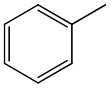
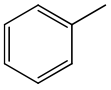
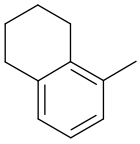
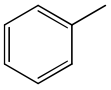
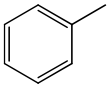
Table 2. Selectivity assays of the 15 compounds for ALR2 and ALR1.

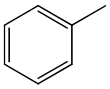
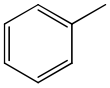
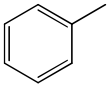
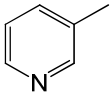
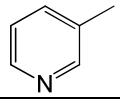
Compd.	% ALR2 inhibition at 50uM [a]	IC_{50} [b] (μM, ALR2)	% ALR1inhibition at 50μM [a]	IC_{50} [b] (μM, ALR1)	Selectivity [c] (ALR1/ALR2)
1	76.17±6.2	10.00±0.59	55.98±2.4	37.93±1.25	3.79
4	53.74±7.6	37.79±0.12	23.61±1.5	ND	-
5	68.80±6.1	31.80±3.38	8.55±1.3	ND	-
7	88.50±1.8	25.05±3.43	57.83±2.4	31.33±0.99	1.25
10	80.70±1.9	10.20±1.83	63.20±5.7	32.82±3.58	3.22
11	68.62±2.2	11.14±0.54	41.13±2.3	107.72±27.66	9.67
13	56.76±4.9	26.55±3.29	7.50±2.2	ND	-
14	94.91±2.4	0.22±0.03	24.20±1.9	ND	-
17	72.30±4.3	15.67±2.76	18.14±5.2	ND	-
18	81.67±1.2	6.30±0.72	82.11±0.8	34.68±1.20	5.50
19	83.61±1.6	10.03±2.33	40.87±5.9	ND	-
22	83.25±2.7	3.65±0.26	5.96±2.5	ND	-
23	70.58±4.5	20.20±1.74	2.62±0.6	ND	-
24	86.72±4.1	4.30±1.20	2.41±0.5	ND	-
25	94.30±1.8	0.89±0.13	41.01±4.3	94.65±15.54	106.35
Epalrestat [d]	92.36±4.3	0.24±0.01	90.48±3.9	2.14±0.13	8.82

[a] % Inhibition values are the mean ± SD of triplicate measurements at 50μM. [b] IC_{50} values shown are the mean ± SD of triplicate measurements. [c] the ratio of ALR1 IC_{50} and ALR2 IC_{50}. [f] Used as positive control. ND: not determined.

Table 3. SAR of acyl-β-phenylalanine and analogues.



Compd.	Source	R_1	R_2	R_3	% of ALR2 inhibition at 50μM [a]	IC_{50} (μM) [b]
14	SYSU-22363S		F	H	94.91±2.4	0.22±0.03
15	SYSU-22410S		OCH_3	H	34.36±3.8	ND
22	SYSU-22433S		H	Cl	83.25±2.7	3.65±0.26
23	SYSU-22439S		H	OCH_3	70.58±4.5	20.20±1.74
24	SYSU-22449S		H	H	86.72±4.1	4.3±1.20
27	SYSU-22364S		F	H	49.49±2.3	ND
28	SYSU-22367S		Cl	H	49.51±4.3	ND
29	SYSU-22370S		F	H	20.56±2.8	ND
30	SYSU-22421S		CH_3	H	45.22±5.2	ND
31	SYSU-22436S		H	CH_3	38.28±1.2	ND

32	SYSU-22454S		H	H	39.85±2.4	ND
33	SYSU-22413S		H	OCH_3	45.5±3.2	ND
34	SYSU-22418S		H	Cl	16.21±1.2	ND
35	SYSU-22368S		OCH_3	H	43.08±5.5	ND
36	SYSU-22437		H	OCH_3	48.94±4.7	ND

[a] Inhibition at 50μM is expressed as the mean ± SD of triplicate measurements. [b] IC_{50} values are the mean ± SD of triplicate measurements. ND: not determined.

Table 4. Cell toxicity of selective ALR2 inhibitors in vitro

Compound	% HEK293 Inhibition at [a]			
	12.5μM	25μM	50μM	100μM
1	16.88±3.7	30.03±3.1	32.92±4.2	34.41±4.3
4	13.63±2.1	14.13±2.7	14.24±3.4	21.12±3.3
5	18.84±2.3	19.97±4.1	21.12±1.8	40.79±3.6
7	15.24±3.5	18.00±2.1	20.24±2.6	22.41±2.9
10	7.39±3.5	11.29±3.3	19.00±4.1	26.29±4.7
11	33.08±3.7	36.58±4.8	50.54±4.3	68.81±4.2
13	7.00±1.2	12.37±2.3	21.00±3.1	52.61±4.3
14	10.46±2.0	14.67±2.4	16.50±3.2	21.05±3.4
17	18.24±2.7	22.00±3.6	25.14±3.8	29.41±2.1
18	14.06±2.0	21.20±1.9	26.69±2.8	30.51±4.0
19	16.76±1.7	24.17±2.4	27.17±3.6	34.63±3.8
22	24.50±2.4	35.03±3.1	39.01±3.9	43.49±4.1
23	15.83±2.1	20.81±2.9	22.61±3.1	41.92±3.4
24	11.35±2.0	17.73±3.1	19.47±2.3	35.87±2.4
25	13.85±1.8	14.69±2.2	19.91±2.7	23.33±1.9
epalrestat[b]	10.32±2.2	12.03±2.9	13.73±3.5	18.28±4.1

[a]HEK293 inhibition is expressed as the mean ± SD of triplicate measurements

[b]Used as positive control

Table of Contents Graphic