PAPER

Check for updates

Decoding selectivity: computational insights into AKR1B1 and AKR1B10 inhibition

Mingyue Liu, [ID] [a] Xiaochun Qin, [a] Jing Li, [a] Yuting Jiang, [b] Junjie Jiang, [a] Jiwei Guo, [a] Hao Xu, [a] Yousen Wang, [a] Hengtai Bi [a] and Zhiliang Wang*[a]

Understanding selectivity mechanisms of inhibitors towards highly homologous proteins is of paramount importance in the design of selective candidates. Human aldo-keto reductases (AKRs) pertain to a superfamily of monomeric oxidoreductases, which serve as NADPH-dependent cytosolic enzymes to catalyze the reduction of carbonyl groups to primary and secondary alcohols using electrons from NADPH. Among AKRs, AKR1B1 is emerging as a promising target for cancer treatment and diabetes, despite its high structural similarity with AKR1B10, which leads to severe adverse events. Therefore, it is crucial to understand the selectivity mechanisms of AKR1B1 and AKR1B10 to discover safe anticancer candidates with optimal therapeutic efficacy. In this study, multiple computational strategies, including sequence alignment, structural comparison, Protein Contacts Atlas analysis, molecular docking, molecular dynamics simulation, MM-GBSA calculation, alanine scanning mutagenesis and pharmacophore modeling analysis were employed to comprehensively understand the selectivity mechanisms of AKR1B1/10 inhibition based on selective inhibitor lidorestat and HAHE. This study would provide substantial evidence in the design of potent and highly selective AKR1B1/10 inhibitors in future.

1. Introduction

Aldo-keto reductases (AKRs) constitute a NADP(H)-dependent and monomeric oxidoreductase superfamily with ubiquitous distribution in human tissues, which mainly located in cytoplasm and catalyze the reduction of carbonyl groups to alcohols for conjugation. To date, 15 human AKRs with over 190 family members have been discovered, which are further divided into six subfamilies (AKR1A, AKR1B, AKR1C, AKR1E, AKR6A, and AKR7A) according to sequence homology and functional similarity.[1–3] They possess a core motif known as the $(\alpha/\beta)8$ barrel, which is alternatively referred to as the TIM barrel due to its association with triose phosphate isomerase, a conserved metabolic enzyme. The $(\alpha/\beta)8$ barrel represents the prevailing fold observed in protein catalysts, accounting for approximately 10% of all documented enzyme structures.[4] Relevant studies have confirmed that AKRs are closely related to various diseases, including diabetic complications, asthma and even cancers.[1,5,6]
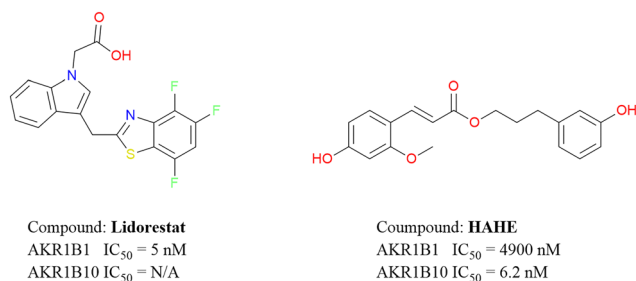
Given the potential role in the progression of diabetes complications and cancers, AKR1B is currently the most extensively studied subfamily, including AKR1B1, AKR1B10 and

AKR1B15.[7,8] AKR1B1, a key rate-limiting enzyme that regulates the polyol pathway of glucose metabolism, facilitates the conversion of glucose to sorbitol using NADPH as a cofactor. Sorbitol is then metabolized to fructose by sorbitol dehydrogenase, thus involving in the occurrence and development of diabetic complications.[1,9] As a well-characterized enzyme, AKR1B10 has a high retinal reductase activity for all-trans-retinaldehyde, which indirectly regulates cell differentiation.[10] It is widely acknowledged that AKR1B1 and AKR1B10 play a pivotal role in the pathology of diabetic complications and are highly expressed in certain cancers, such as breast, ovarian and lung cancer.[11–13] The application of AKR1B1 or AKR1B10 inhibitors has been considered as a predominant therapeutic strategy to treat diabetic complications, cancers and other diseases. Structurally, AKR1B1 and AKR1B10 not only share a 70.6% amino acid sequence identity but also overlap in substrate specificities for aliphatic and aromatic aldehydes.[10] This is the reason why AKR1B1 inhibitors can effectively bind to AKR1B10 proteins.

Despite rigorous screening and optimization of most AKRs inhibitor candidates entering the preclinical stage, approximately 60% of drug candidates do not progress beyond phase I, II, and III clinical trials and are even withdrawn due to poor selectivity, resulting in elevated toxicity, side effects and potentially unmanageable outcomes.[14] One example is tolrestat, a typical AKR1B1 inhibitor developed by Ryerst Inc. and first marketed in Ireland for the treatment of adult diabetic

[a] Department of Drug Clinical Research Center, The First Affiliated Hospital of Shandong Second Medical University, Weifang 261000, China. E-mail: wangzhiliang1977@126.com

[b] School of Pharmacy, Harbin Medical University, Harbin 150081, China

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 9295–9308 | **9295**

Fig. 1 Structures and inhibitory activity of the highly selective AKR1B1 and AKR1B10 inhibitors.

retinopathy in 1989, which has been withdrawn because of severe hepatotoxicity, which is mainly attributed to its lower inhibitory selectivity against AKR1B1 over AKR1B10.[15] Up to now, epalrestat serves as the only AKR1B1/10 inhibitor commercially available for the treatment of diabetic complications.[16,17] Therefore, it is crucial to understand the reasons for the poor selectivity of AKR1B1 and AKR1B10 to develop AKRs selective inhibitors.

Here, our study focused on the structural properties of selective AKR1B1 and AKR1B10 inhibitors within the coenzyme binding region, catalytic region and C-terminal free ring structure of AKR1B1/10 proteins. We selected two representative inhibitors: lidorestat,[18] a known selective inhibitor of AKR1B1, for treatment of chronic diabetic complications and HAHE,[19] a selective inhibitor of AKR1B10 that is 790-fold more potent than AKR1B1, significantly suppressed farnesal metabolism and cell proliferation in AKR1B10-overexpressing cells (Fig. 1). To thoroughly examine the correlation between the active sites of AKR1B1 and AKR1B10 and their selective inhibitors, a series of silico-based approaches were employed in a sequential manner.[20] These approaches encompassed sequence alignment, molecular docking, molecular dynamics simulation, MM-GBSA calculation,[21] alanine scanning mutagenesis[22] and pharmacophore modeling analysis. Our research sheds light on the selectivity mechanism of AKR1B1/10 inhibitors and instills great assurance for future advancements in the design and optimization of selective inhibitors targeting AKR1B1 or AKR1B10.

2. Methods

2.1. Sequence alignment and structure superposition

To determine the structural similarities and differences, the Discovery Studio 3.0 software package was used to align the genomic full-length sequence of human AKR1B1 (Uniprot code P15121) and AKR1B10 (Uniprot code O60218) downloaded in FASTA format from Uniprot (https://www.uniprot.org/).[23] Meanwhile, the human AKR1B1-NADP⁺-lidorestat complex (PDB code:1Z3N)[18] and AKR1B10-NADP⁺-tolrestat complex (PDB code:1ZUA)[10] were obtained from RSCB Protein Data Bank (https://www.rcsb.org)[24] and were used for the structural superposition. Each co-crystallized ligand was anchored in the binding pocket of the receptors (AKR1B1 and AKR1B10) to

automatically align using the "Superimpose Proteins" module in the Discovery Studio 3.0 software (Accelrys, San Diego, CA, USA).

2.2. Protein Contacts Atlas analysis

The online server at https://www.mrc-lmb.cam.ac.uk/pca/ was used to import the crystal structure of AKR1B1 and AKR1B10 for Protein Contacts Atlas analysis.[25] This analysis aimed to examine binding characteristics and understand protein–ligand interactions of AKR1B1 and AKR1B10 through residue-residue interaction networks.

2.3. Protein and ligand preparation

The crystal structures of human AKR1B1/NADP⁺/lidorestat complex (PDB code: 1Z3N) and AKR1B10/NADP⁺/tolrestat complex (PDB code:1ZUA) were imported into Maestro version 13.5 from the Schrödinger package (Schrödinger, LLC, New York, NY, 2023)[26] and then prepared by the Protein Preparation Workflow[27] module, which removed the original hydrogen atoms, added hydrogen atoms, filled in the missing side chains and loops using the Prime module.[28] Subsequently, the options for "optimal H-bonds assignment, removal of water and protein minimization under the OPLS-4 force field" were selected to relieve any clashes. Additionally, lidorestat and HAHE were drawn in "sdf" format using ChemDraw19.0 Ultra and optimized using the OPLS-4 force field[29] to generate low-energy conformers automatically using the Ligprep module[30] of the Schrödinger suite (Schrödinger, LLC, New York, NY, 2023) for subsequent molecular docking.

2.4. Molecular docking

Molecular docking was performed using Glide 9.7[31,32] embedded in Maestro version 13.5 from the Schrödinger package. A grid box was constructed to define the active site using the Receptor Grid Generation module, with the center of the co-crystal ligand serving as the reference point. In order to ensure the docking results accurately and effectively, the co-crystal ligands were re-docked to the binding pockets of AKR1B1 (PDB code:1Z3N) and AKR1B10 (PDB code: 1ZUA) in the "Extra Precision" (XP) mode to further compute Root Mean Square Deviation (RMSD) values. Then, all compounds were docked flexibly in the same manner to predict all possible binding conformations, which were sorted by the docking score, glide score and glide emodel using the Ligand Docking module.[33] Finally, PYMOL 2.5 was used to draw the optimal binding conformations, which were then selected for the subsequent molecular dynamics simulation.

2.5. Molecular dynamics (MD) simulation

To further optimize the binding mode of AKR1B1/10-NADP⁺-lidorestat/HAHE complexes, we performed conventional molecular dynamics simulations using the Desmond program.[34] To parameterize the AKR1B1/10 protein and small molecules (lidorestat and HAHE), the OPLS_2005 force field was utilized, with the water solvent being modeled using the TIP3P model. The protein–small molecule complex was placed in a cubic

9296 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 9295–9308

This journal is © the Owner Societies 2024

water box and solvated. To neutralize the charge of the system, chloride and sodium ions were added at a concentration of 0.150 M. Initially, the system's energy was minimized by employing the steepest descent minimization technique for a total of 50 000 steps. Subsequently, the positions of large atoms were limited during NVT and NPT equilibration for an extra 50 000 steps. The temperature of the system was kept at 300 Kelvin, while the pressure of the system was maintained at 1 bar. Following the completion of the two equilibration phases, a simulation without any restrictions was conducted for a duration of 100 nanoseconds. During the 100 ns MD simulation, RMSD and Root Mean Square Fluctuation (RMSF) were monitored. Finally, the protein–ligand interactions were analyzed and dynamic trajectory animations were generated using Maestro version 13.5.

2.6. Molecular mechanics/generalized Born surface area (MM/GBSA) calculation and energy decomposition calculation

According to the aforementioned AKR1B1/10 and their selective inhibitor complexes, the binding free energy of the protein–ligand complex was calculated using the MM-GBSA method from Schrödinger Maestro version 13.5 and AMBER version 18.[35] First, the equilibrated coordinates of complexes were obtained through MD simulations performed using AMBER version 18 with the ff14SB force field. Force field parameters of the ligand were generated using the general AMBER force field GAFF2 through the Antechamber program. The hydrogen atoms were positioned using the LEaP module, and each complex was placed within a cubic box containing TIP3P water molecules with a solute-wall distance of 10 Å. Counter ions were introduced to maintain system neutrality. The system underwent energy minimization through 2500 steps using the steep descent algorithm followed by 2500 steps using the conjugate gradient algorithm, with a nonbonded cutoff of 10 Å. Then, MD simulation was conducted involving gradual heating, density equilibration, equilibration and 100 ns production procedures. Each complex was subjected to three repetitions of the simulation using the equilibrated part of the MD trajectories. Totally, 100 snapshots were extracted from the equilibrium trajectory for MM-GBSA free energy calculation according to the following equations:[36]

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand}) \tag{1}$$

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \tag{2}$$

$$\Delta E_{MM} = \Delta E_{ele} + \Delta E_{vdW} + \Delta E_{int} \tag{3}$$

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA} \tag{4}$$

where ΔG_{blind} represents the total binding free energy of protein–ligand complexes, while ΔE_{MM} represents the gas-phase interaction energy, which is decomposed into electrostatic interaction energy (ΔE_{ele}), van der Waals interaction energy (ΔE_{vdW}) and internal energy of the system (ΔE_{int}). $-T\Delta S$ represents the change in ligand-binding conformational entropy where T and S denote the absolute temperature and entropy, respectively. Typically, $-T\Delta S$ is calculated through

normal mode analysis, but it is ignored in this study due to the significant time consumption, uncertainty of the contributions to the total free energy and the large margin of error. In addition, ΔG_{GB} and ΔG_{SA} denote the polar and non-polar contributions of solvation-free energy (ΔG_{sol}), respectively. Besides, ΔG_{GB} is often calculated based on the Generalized Born model (GB), while ΔG_{SA} is calculated based on surface tension and solvent-accessible surface area. Additionally, a per-residue energy decomposition analysis was performed to determine the energy contribution of each residue to the overall binding energy based on the above same snapshots.

2.7. Alanine scanning mutagenesis analysis

In order to identify the effect of specific amino acid residues on the AKR1B1/10-lidorestat/HAHE complexes surface, alanine scanning mutagenesis analysis[22,37] was carried out to calculate the variation in binding free energy ($\Delta\Delta G_{bind}$) between the mutant and the wild-type system. Typically, the contribution of alanine–mutant's binding free energy is negligible. Therefore, the gap in the binding free energy between the mutant and wild-type system corresponds to the involvement of mutant residues in the overall binding free energy. Please refer to the formula below for details.

$$\Delta\Delta G_{bind}^{wild \to mutant} = \Delta\Delta G_{bind}^{mutant} - \Delta\Delta G_{bind}^{wild} \tag{5}$$

where $\Delta\Delta G_{bind}^{mutant}$ and $\Delta\Delta G_{bind}^{wild}$ represent the binding free energy of the wild-type system and mutant system after residue mutation to alanine, respectively.
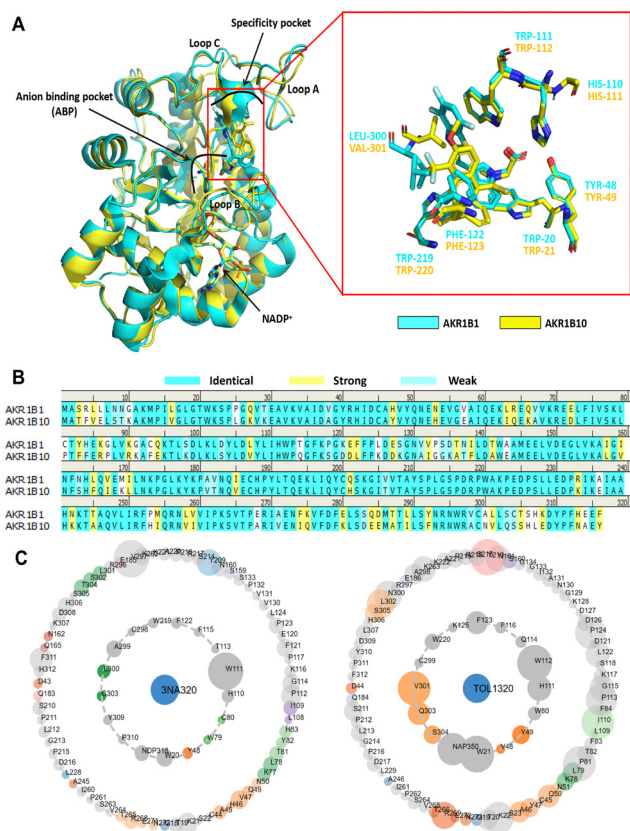
2.8. Structure-based 3D pharmacophore models

Based on the optimal binding mode of AKR1B1/10-lidorestat/HAHE complexes, the three-dimensional (3D) pharmacophore models were automatically generated by defining the active site of lidorestat and HAHE using the LigandScout 4.4.8 software.[38] Some chemical features, including hydrogen bond donor/acceptor, aromatic ring center and hydrophobic group, were observed in the 3D pharmacophore models. Leave all other parameters as the default values throughout the entire process.

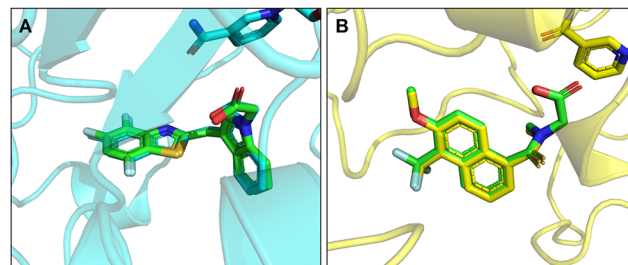## 3. Results and discussion

3.1. Comparison of AKR1B1/10 structures and sequences

To gain preliminary insights into the selective mechanism of AKR1B1/10 inhibition, the comparison of the structural similarities and differences between AKR1B1 and AKR1B10 was performed through sequence alignment and structure superposition. The crystal structures of AKR1B1 (PDB code: 1Z3N) and AKR1B10 (PDB code: 1ZUA) were subsequently chosen for further analysis. As shown in Fig. 2(A) and (B) AKR1B1 and AKR1B10 possessed a high degree of sequence homology with the sequence identity of 70.6% and similarity up to 85.8%. Meanwhile, some key residues involving in catalytic residues TYR48, HIS110 (AKR1B1 numbering) and TRP111 in the active site of the AKR1B1 holoenzyme structure overlapped tightly

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 9295–9308 | **9297**

Fig. 2 Comparison of AKR1B1 and AKR1B10 in structure and sequence. (A) Structure superposition of the AKR1B1-NADP$^+$-lidorestat complex (PDB code: 1Z3N, blue ribbon) and AKR1B10-NADP$^+$-tolrestat complex (PDB code: 1ZUA, yellow ribbon). The red boxes represent key amino acids formed by ligands that bind to AKR1B1 and AKR1B10, which are shown in blue and yellow sticks, respectively. (B) Sequence alignment of AKR1B1 and AKR1B10. The blue characters indicate the identical residues of AKR1B1 and AKR1B10, the yellow and white characters indicate similar and non-matching residues, respectively. (C) The asteroid plot formed by Protein Contact Atlas based on the AKR1B1 (left) and AKR1B10 (right) co-crystal structure. The ligand, represented as the central node and highlighted in blue, is surrounded by residues that are color-coded according to their secondary structures. The inner shell residues refer to the immediate neighbors that directly interact with the ligand molecule, whereas the outer shell residues engage in indirect interactions. The size of the circular nodes corresponds to the number of contact residues established with the ligand.

with these in the AKR1B10–tolrestat complex structure, which were composed of TYR49, HIS111 (AKR1B10 numbering) and TRP112. Of note, these key residues define a geometrically rigid "anion binding pocket (ABP)" with the nicotinamide moiety of NADP$^+$, which was occupied by the negatively charged groups of AKR1B1/10 inhibitors. Furthermore, the Protein Contacts Atlas analysis was performed to examine the important non-covalent interactions between AKR1B1 and AKR1B10 proteins and their co-ligands. The results further confirmed the significance of these proposed crucial residues (Fig. 2(C)). Overall, the significant resemblance between the structures of AKR1B1 and AKR1B10 poses a challenge in the development of highly selective small molecule inhibitors.

Fig. 3 Comparison of the predicted and experimental poses for (A) AKR1B1, RMSD = 0.285 Å (PDB code: 1Z3N) and (B) AKR1B10, RMSD = 0.220 Å (PDB code: 1ZUA). The green sticks represent the experimental pose extracted from the X-ray structure, whereas the blue and yellow sticks represent the docked poses.

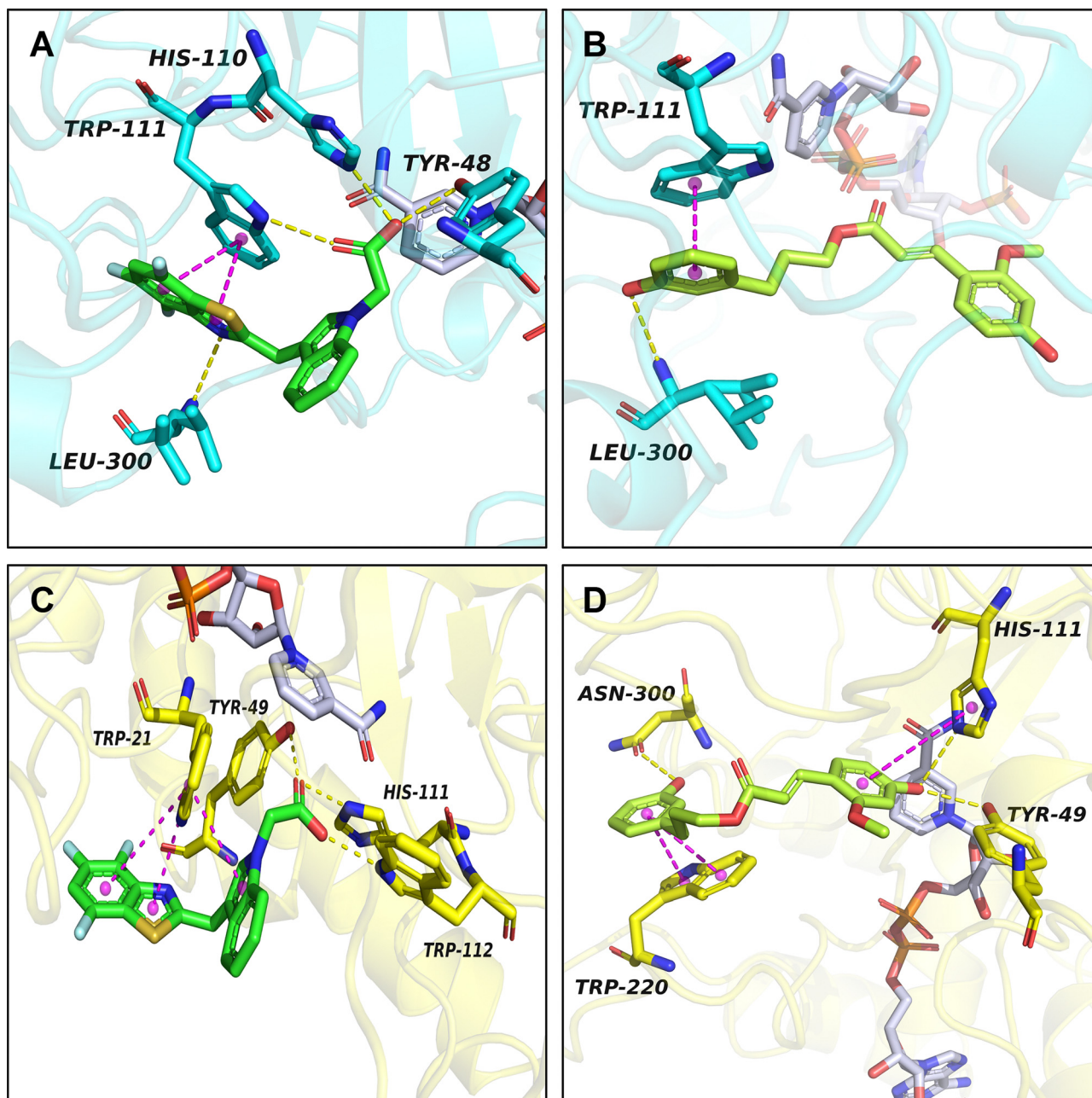## 3.2. Comparison of binding modes against AKR1B1/10 inhibitors through molecular docking

Molecular docking is a commonly employed method for evaluating the interaction patterns and binding affinities between receptors and ligands. To elucidate the underlying structural reasons for the selectivity mechanism of AKR1B1/10 inhibition, the native ligands from AKR1B1 and AKR1B10 were initially re-docked into AKR1B1-NADP$^+$ and AKR1B10-NADP$^+$ complexes with the optimized docking parameters to ensure the reliability of molecular docking results. The results showed that the optimal conformation of the co-crystalline ligands was successfully attached to the crystal structure of AKR1B1 and AKR1B10, with RMSD values of 0.285 Å and 0.220 Å, respectively, suggesting that the docking protocol is reliable and accurate (Fig. 3).

Then, lidorestat and HAHE were docked into the catalytic region of AKR1B1 and AKR1B10 using the same parameters as mentioned above. As indicated in Table 1, the docking score of lidorestat and HAHE against AKR1B/10 proteins was compatible with their biological activity tendency. For example, the selective AKR1B1 inhibitor lidorestat displayed a better docking score of −12.402 kcal mol^{-1} towards 1Z3N than the selective AKR1B10 inhibitor HAHE (−8.292 kcal mol^{-1}), indicating that lidorestat binding to AKR1B1 was superior to HAHE. In contrast, lidorestat exerted an opposite effect with a docking score of −3.940 kcal mol^{-1} toward 1ZUA compared to HAHE (−9.311 kcal mol^{-1}). In addition, the Glide emodel was devoted to picking the "best" pose of a ligand (pose selection) with a more significant weighting of the force field components (electrostatic and van der Waals energies), and then ranks these best poses against one another with GlideScore, which is considered as the empirical scoring function that estimates

Table 1 Glide docking scores of lidorestat and HAHE against AKR1B1/10 proteins

Complex	Docking score (kcal mol^{-1})	Glide score (kcal mol^{-1})	Glide emodel (kcal mol^{-1})
1Z3N/lidorestat	−12.402	−12.402	−106.697
1Z3N/HAHE	−8.292	−8.292	−70.150
1ZUA/lidorestat	−3.940	−3.940	−61.826
1ZUA/HAHE	−9.311	−9.311	−62.424

9298 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 9295–9308

This journal is © the Owner Societies 2024

Fig. 4 Predicted binding patterns of lidorestat and HAHE against AKR1B1-NADP$^+$ (PDB code: 1Z3N, blue ribbon) and AKR1B10-NADP$^+$ complex (PDB code: 1ZUA, yellow ribbon). (A) AKR1B1 with lidorestat. (B) AKR1B1 with HAHE. (C) AKR1B10 with lidorestat. (D) AKR1B10 with HAHE. Lidorestat and HAHE were displayed as the green and chartreuse sticks, respectively. The gray sticks represent the coenzyme named as NADP$^+$. The yellow dashed line represents the hydrogen bonds and the magenta dashed line represents the π−π stacking interaction.

the binding free energy of the ligand.[31–33] It was obvious that the Glide emodel further confirmed the above results (Table 1). Although these results demonstrated that lidorestat exerted a stronger binding affinity for AKR1B1 compared to AKR1B10, while HAHE showed a more favorable impact on AKR1B10, experiments are still needed for further verification.

As demonstrated in Fig. 4 and 5, the oxygen atoms on the carboxyl group of lidorestat formed strong hydrogen bond interactions with catalytic residues TYR48, HIS110 and key residue TRP111. These interactions effectively anchored

lidorestat within the active site of AKR1B1, thereby significantly enhancing the enzymatic activity of AKR1B1. It is worth noting that the "specificity pocket" responsible for binding AKR1B1 consists of TRP111, THR113 and PHE122 from loop A, as well as CYS298 to SER302 and TYR309 from loop C, particularly in close proximity to LEU300, which exhibits considerable flexibility. The opening of this pocket is typically induced by the hydrophobic group of the inhibitor and plays a vital role in AKR1B1 selectivity. Obviously, in the AKR1B1-NADP$^+$-lidorestat complex, the specificity pocket of AKR1B1 exhibits an

This journal is © the Owner Societies 2024

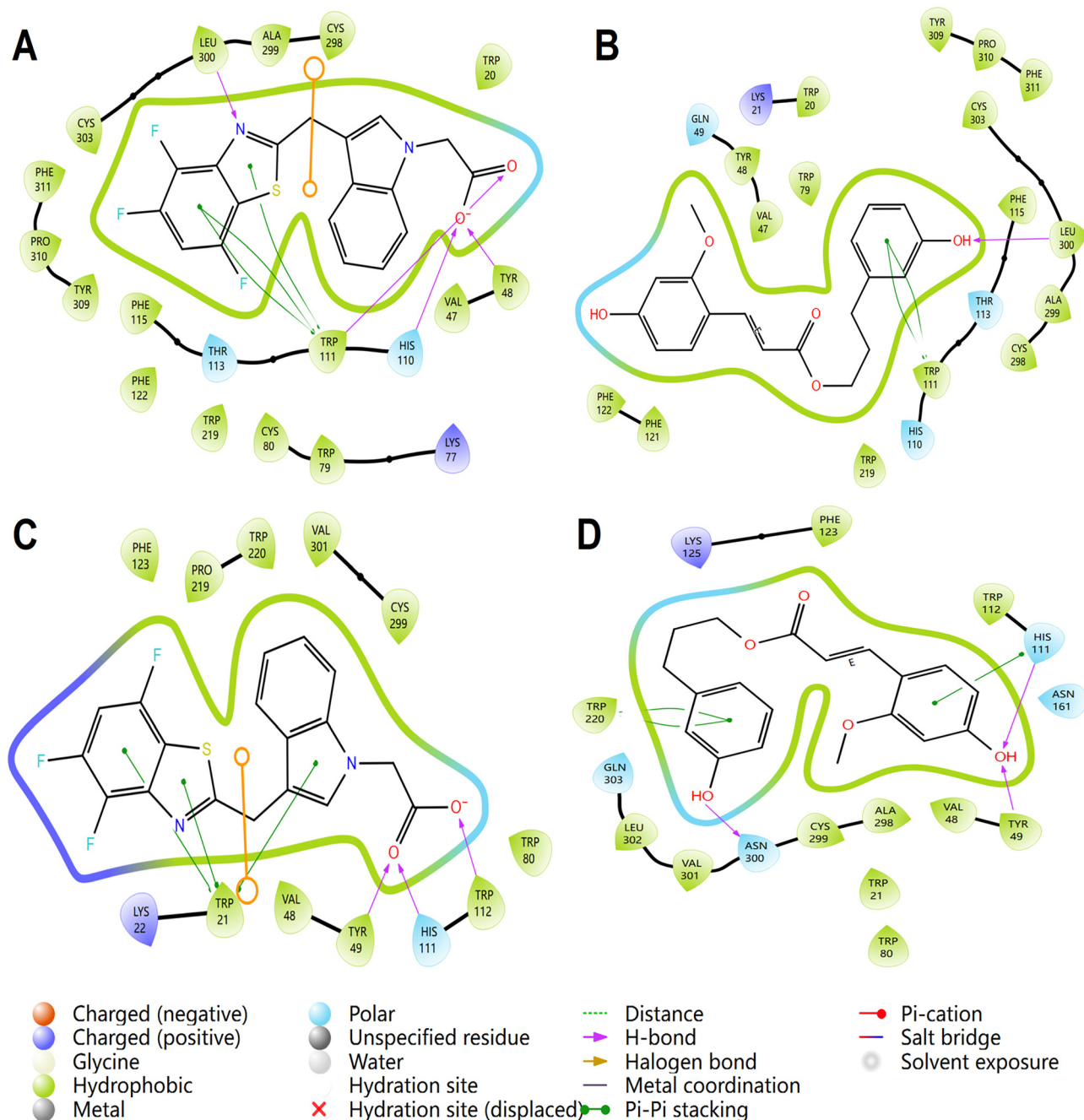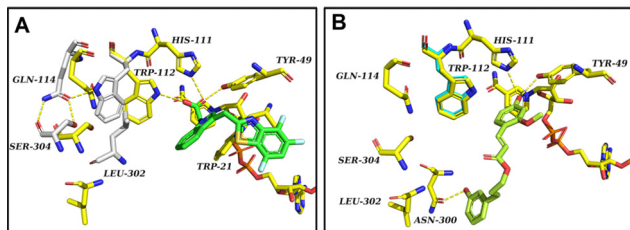*Phys. Chem. Chem. Phys.*, 2024, **26**, 9295–9308 | **9299**

Fig. 5 Two-dimensional binding patterns of lidorestat and HAHE against AKR1B1 (PDB code: 1Z3N, blue ribbon) and AKR1B10 (PDB code: 1ZUA, yellow ribbon). (A) AKR1B1 with lidorestat. (B) AKR1B1 with HAHE. (C) AKR1B10 with lidorestat. (D) AKR1B10 with HAHE.

additional hydrogen bond between the nitrogen atoms on the benzothiazole group of lidorestat and the amide N–H of LEU300, alongside an almost flawless π–π stacking interaction involving the benzothiazole side chain and the indole moiety of TRP111, resulting in a transition to an open state. These interactions contribute to the high selectivity of lidorestat against AKR1B1. In contrast, HAHE only forms a hydrogen bond with LEU300 and π–π stacking with TRP111 of AKR1B1, which does not impact the activity of AKR1B1. These results suggest that selective AKR1B1 inhibition can potentially be

achieved by elongating the side chain of AKR1B1 inhibitors with a bulky group to occupy a larger interspace within the hydrophobic sub-pocket of AKR1B1. This would also improve the capacity to establish robust hydrogen bond connections with the catalytic residues TYR48, HIS110, and the crucial residue TRP111.

As for AKR1B10, the hydroxyl oxygen atoms on the benzene ring of HAHE and the oxygen atoms on the carboxyl group of lidorestat occupied the anionic site and formed critical hydrogen bonds with catalytic amino acids TYR49 and HIS111. In
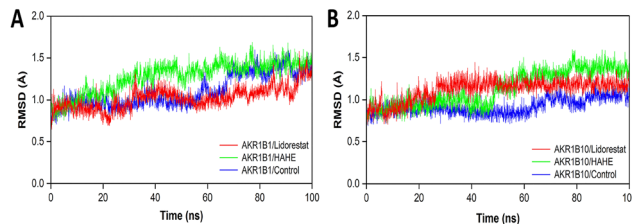
9300 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 9295–9308

This journal is © the Owner Societies 2024

Fig. 6 Change in the TRP112 side-chain conformation in AKR1B1 and AK1RB10. (A) and (B) The superposition of the holo-AKR1B10 (gray sticks) active site structure (PDB code: 4GQG) with that of the AKR1B10-NADP$^+$-lidorestat docked complex (yellow and green stick) and the holo-AKR1B1 (blue stick) active site structure (PDB code: 1Z3N) with that of the AKR1B10-NADP$^+$-HAHE docked complex (yellow and chartreuse stick), respectively. Hydrogen bonds are shown as dashed yellow lines.

Fig. 7 RMSD of protein backbone atoms monitored throughout the 100 ns molecular dynamics simulations. (A) AKR1B1/lidorestat (red line), AKR1B1/HAHE (green line) and AKR1B1/Control (blue line) complexes. (B) AKR1B10/lidorestat (red line), AKR1B10/HAHE (green line) and AKR1B10/Control (blue line) complexes.

addition, the oxygen atoms on the side chain benzene ring of HAHE formed additional hydrogen bonds with ASN300 and engaged in π–π stacking with HIS111 and TRP220, thereby further augmenting the activity of AKR1B10 compared to lidorestat. Interestingly, lidorestat also formed hydrogen bonds with TRP112 of AKR1B10. Zhang *et al.* have identified a surprising TRP112 native conformation stabilized by a specific GLN114-centered hydrogen bond network in the AKR1B10 holoenzyme, which could play a critical role in the determination of AKR1B1/10 inhibitor selectivity. They have confirmed that the highly selectivity of AKR1B10 inhibitor flufenamic acid was due to the steric clash that TRP111 in AKR1B1 (always in flipped position) would have with the benzoic acid moiety of the inhibitor, and that the native TRP112 position is avoided in AKR1B10.[39] Subsequently, we further compared the superposition of the holo-AKR1B10 active site structure with that of the AKR1B10-NADP$^+$-lidorestat complex and the holo-AKR1B1 active site structure with that of AKR1B10-NADP$^+$-HAHE, respectively. As shown in Fig. 6(A), lidorestat can induce an "AKR1B1-like" active site in AKR1B10, whereas it is incapable of opening the "specificity pocket" in AKR1B10, which was in accordance with tolrestat in a previous study, indicating that the activity of lidorestat towards AKR1B10 was equivalent to tolrestat and inferior to HAHE. Similarly, AKR1B10 selective inhibitor HAHE caused a flip of TRP112, making it also induce an "AKR1B1-like" active site in AKR1B10. However, this result is in contrast to that of flufenamic acid in AKR1B10. On the other hand, the benzene ring of HAHE adjacent to the anionic site of AKR1B1 established π–π stacking and hydrophobic interaction with the TRP111 indole ring, elucidating the weak inhibition of HAHE to AKR1B1 than AKR1B10 (Fig. 6(B)).

3.3. Comparison of dynamics behaviors between AKR1B1 and AKR1B10

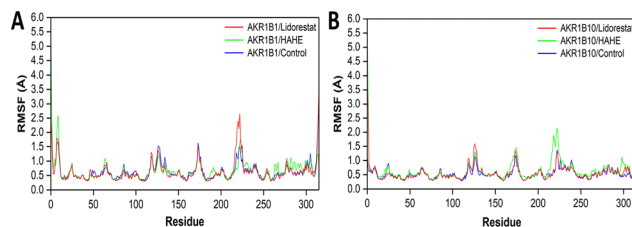### 3.3.1. Stability of dynamics trajectory from RMSD analysis.
In order to elucidate the dynamic behavior of the residues located at the active sites of AKR1B1/10, the complexes' crystal structures were subjected to molecular dynamics simulations in a realistic solution environment. The apo structure of both AKR1B1 and AKR1B10 was selected as the control. The system's

stability was assessed by monitoring the fluctuation of the RMSD value, which serves as an indicator of the conformational changes occurring within the complex throughout the course of the MD simulation. Generally, it is acceptable for the RMSD value to fluctuate within a range of 3. Notably, the RMSD values of the α-carbon (Cα) atoms within all the complexes ultimately converged towards stability in the final 20 ns (Fig. 7), indicating that the structures of each complex reached stable states throughout the simulations.

3.3.2. Structural flexibility from RMSF analysis.
Subsequently, the RMSF diagram of Cα atoms was utilized to analyze the fluctuations in each complex system and compared to the apo structure, as depicted in Fig. 8, lidorestat and HAHE exerted comparable RMSF dynamics profiles in the system when interacting with AKR1B1/10, indicating that both inhibitors possess comparable binding modes at the binding pocket of AKR1B1/10. Meanwhile, slight RMSF peaks observed in residue TRP219 of AKR1B1 and TRP220 of AKR1B10 suggest their involvement in stable interactions with the inhibitors. Furthermore, it is observed that the AKR1B10 structure exhibited smaller RMSF fluctuations compared to AKR1B1, potentially due to hydrophobic interactions among residues within the AKR1B1 binding pocket. In particular, the flexibility of the loop region, as well as the C- and N-terminus, was evident through the considerable fluctuation observed in the RMSF curves. Furthermore, the flexibility of residues in the binding pocket of AKR1B10 was also restricted by hydrogen bonds and water bridges.

Fig. 8 RMSF plot from MD simulation. (A) RMSF of each residue of the protein for the complex obtained from AKR1B1/lidorestat (red line), AKR1B1/HAHE (green line) and AKR1B1/control (blue line) complexes obtained from 100 ns MD simulation and (B) the RMSF maps of AKR1B10/lidorestat (red line), AKR1B10/HAHE (green line) and AKR1B10/control (blue line) complexes in the entire 100 ns MD simulations.
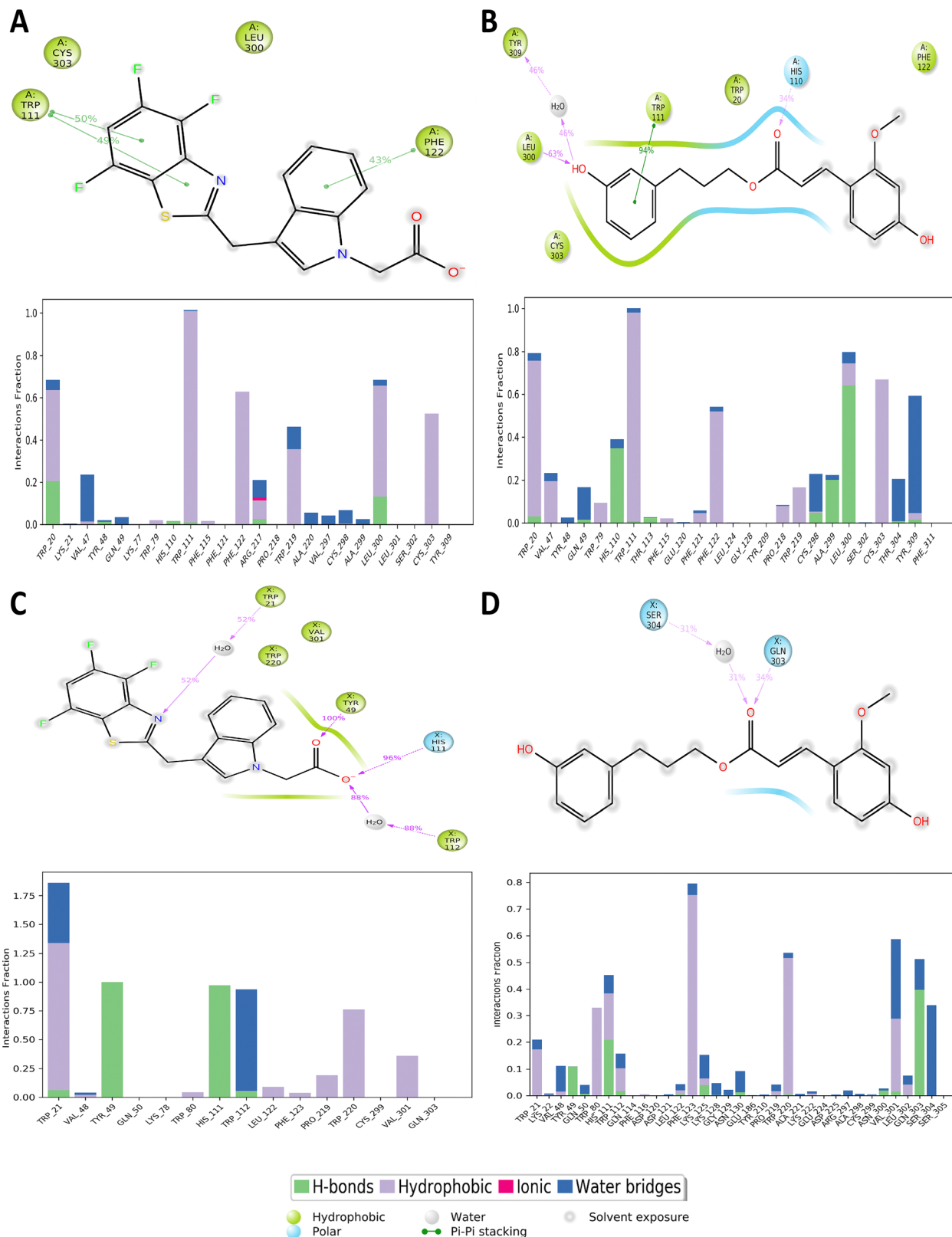
This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 9295–9308 | **9301**

Fig. 9 Intermolecular interaction analysis of protein–ligand interactions obtained from the last frame of the MD simulation trajectory for complexes of AKR1B1/10. (A) AKR1B1/lidorestat. (B) AKR1B1/HAHE. (C) AKR1B10//lidorestat. (D) AKR1B10/HAHE.

9302 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 9295–9308

This journal is © the Owner Societies 2024

3.3.3. Intermolecular interaction analysis of AKR1B1/10 inhibitors from MD simulation. Intermolecular interaction analysis was conducted to further elucidate the unique action mechanisms among lidorestat, HAHE and AKR1B1/10 proteins. During the MD simulation, this study monitored three primary forms of protein–ligand connections, which included hydrogen bonding, hydrophobic bonding, and water bridges. These interactions exceeding 1% were recorded for analysis. As demonstrated in Fig. 9(A), hydrophobic interactions were found to be the primary contacts between lidorestat and the AKR1B1 residues TRP111 and PHE122, contributing to 50%, 49%, and 43% of the overall interaction, with no hydrogen bonds found. Among them, TRP111 directly formed π–π stacking with the benzothiazole groups of lidorestat, aligning with the findings from molecular docking experiments. In the AKR1B1/HAHE complex, HAHE also formed π–π stacking with TRP111, accounting for 94%. Besides, hydrogen bonds were also formed between HAHE and LEU300 (63%) and HIS110 (34%), as well as a water bridge with TYR309 (46%). In Brief, in order to improve the selectivity of inhibitors against AKR1B1, one can focus on modifying the structure of current inhibitors to strengthen their binding with TRP111 and PHE122.

Considering the AKR1B10/lidorestat complex in Fig. 9(C), lidorestat formed hydrogen bonds with TYR49 (100%) and HIS111 (96%) and two water bridge interactions with TRP21 (52%) and TRP112 (88%) were observed. In contrast, the key amino acids responsible for the interaction between the AKR1B10 protein and HAHE were HIS111, PHE123, TRP220, VAL301, and GLN303. These amino acids primarily engaged in hydrophobic and water bridge interactions, as depicted in Fig. 9(D). Strangely, HAHE exhibited a hydrogen bond exclusively with GLN303 (34%) and a water bridge was observed with SER304 (31%), deviating from the molecular docking findings of the AKR1B10/HAHE complex.
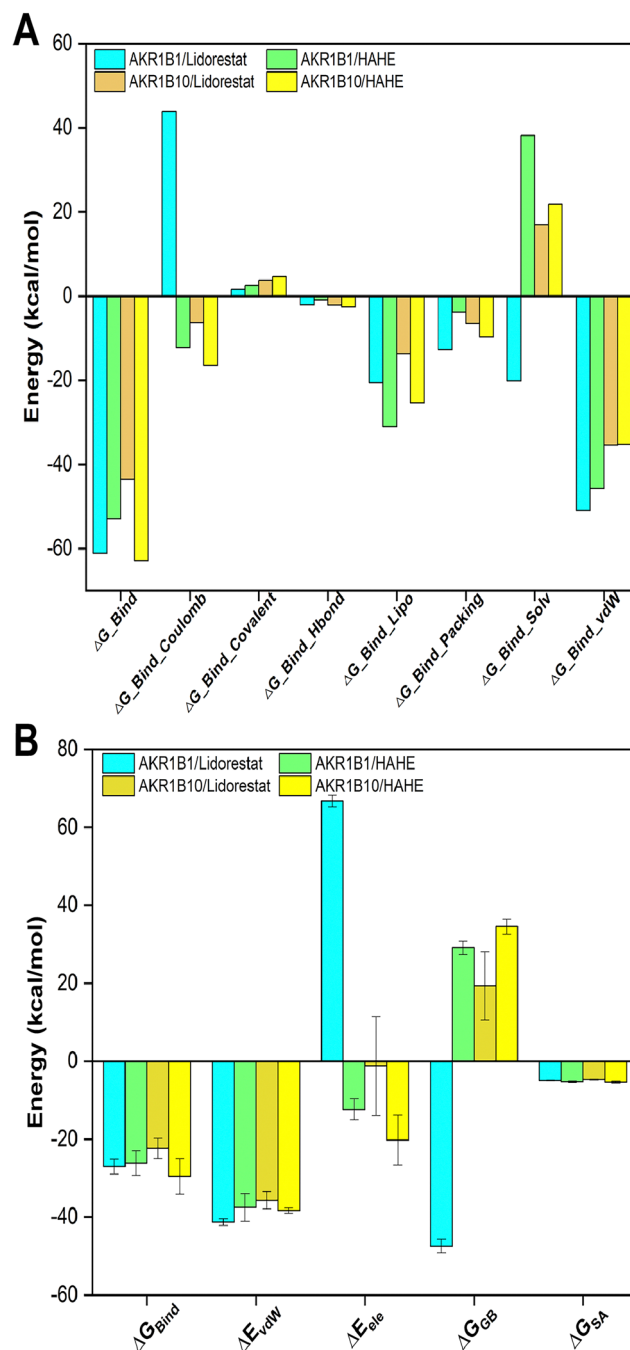
3.4. Calculation of the AKR1B1/10 composite structures' binding free energy

To assess the binding affinities of the lidorestat, HAHE and AKR1B1/10 proteins, the Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) method was used to calculate the binding free energy, which included ΔG_Bind, ΔG_Bind_Coulomb, ΔG_Bind_Covalent, ΔG_Bind_Hbond, ΔG_Bind_Lipo, ΔG_Bind_packing, ΔG_Bind_Solv GB and ΔG_Bind_vdW from Schrödinger Maestro version 13.5. The AKR1B1/10-lidorestat/HAHE binding models were derived from the result of the MD simulation. As demonstrated in Table 2, the MM-GBSA calculations yielded binding free energies of -61.12 kcal mol^{-1} and -62.99 kcal mol^{-1} for the AKR1B1/lidorestat and AKR1B10/HAHE complexes, respectively. These values aligned with the experimentally-measured affinities, suggesting that lidorestat

Fig. 10 The spectrum of intermolecular interaction between protein–ligand complexes determined from MM-GBSA calculation based on the Schrödinger Maestro version 13.5 (A) and the AMBER version 18 (B), respectively.

Table 2 Complexes binding free energy calculated (kcal mol^{-1}) using the MM-GBSA algorithm from Schrödinger Maestro version 13.5

Energy	AKR1B1 (1Z3N, kcal mol^{-1})		AKR1B10 (1ZUA, kcal mol^{-1})	
	Lidorestat	HAHE	Lidorestat	HAHE
ΔG_Bind	−61.12	−52.88	−43.56	−62.99
ΔG_Bind_Coulomb	43.90	−12.21	−6.36	−16.46
ΔG_Bind_Covalent	1.59	2.58	3.81	4.65
ΔG_Bind_Hbond	−2.09	−0.93	−2.18	−2.56
ΔG_Bind_Lipo	−20.64	−31.02	−13.80	−25.47
ΔG_Bind_packing	−12.76	−3.78	−6.55	−9.73
ΔG_Bind_Solv GB	−20.15	38.13	16.88	21.81
ΔG_Bind_vdW	−50.97	−45.65	−35.36	−35.23

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 9295–9308 | 9303

exhibited stronger binding affinities towards AKR1B1. Nevertheless, lidorestat and HAHE exhibited an inferior binding affinity towards AKR1B10 (-43.56 kcal mol^{-1}) and AKR1B1 (-52.88 kcal mol^{-1}), respectively. Furthermore, it is worth noting that the lipophilic energy (-20.64 kcal mol^{-1} for AKR1B1/lidorestat and -25.47 kcal mol^{-1} for AKR1B10/HAHE) and the van der Waals energy (-50.97 kcal/mol for AKR1B1/lidorestat and -30.23 kcal mol^{-1} for AKR1B10/HAHE) contributed the most. In particular, the AKR1B1/lidorestat complex exhibited a coulomb energy of 43.90 kcal mol^{-1}, which was opposed to the other three complexes. This result probably demonstrated that the binding process seems to be stabilized by reducing the polar interaction between the ligand and the receptor. By thoroughly analyzing the data on hydrogen bond energy provided in Table 2 and Fig. 10(A), one can conclude

that hydrogen bonds primarily aided in the identification of ligands by proteins.

To explore residues' contributions in above complexes, four groups of 100 ns equilibrium MD were performed using AMBER version 18 and then MM-GBSA binding free energy was further calculated. As shown in Fig. 10(B), the AKR1B1/lidorestat complex formed stronger van der Waals interaction than the AKR1B10/HAHE complex, whereas the AKR1B10/HAHE complex established stronger electrostatic forces than the AKR1B10/lidorestat complex. It is worth noting that van der Waals energy is the major contributor in the entire binding free energy. The discovery provides additional evidence for the previous hypothesis regarding the connection between the hydrophobic cavity of AKR1B1/10 and the heightened molecular specificity observed in the analysis of their interaction.
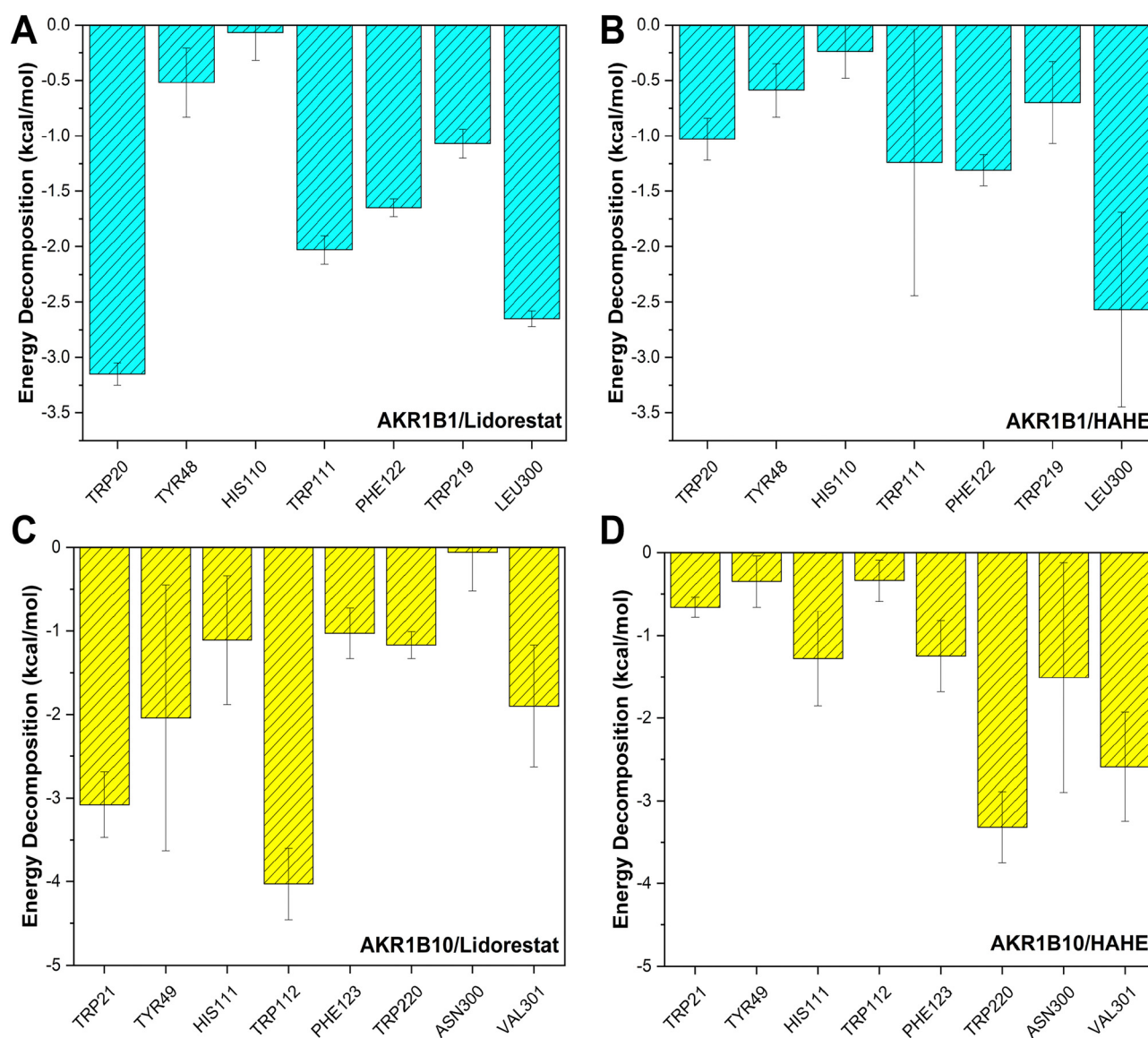

Fig. 11 Decomposition of the MM-GBSA binding energy per residue in the binding pocket of each complex. (A) MM-GBSA energy decomposition scheme of the AKR1B1/lidorestat complex. (B) MM-GBSA energy decomposition scheme of the AKR1B1/HAHE complex. (C) MM-GBSA energy decomposition scheme of the AKR1B10/lidorestat complex. (D) MM-GBSA energy decomposition scheme of the AKR1B10/HAHE complex.

9304 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 9295–9308

This journal is © the Owner Societies 2024

Furthermore, the per-residue energy decomposition analysis using the MM-GBSA method was applied to evaluate the energy contributions of each residue toward the overall binding free energy in the activation region of the AKR1B1-NADP$^+$-lidorestat and AKR1B10-NADP$^+$-HAHE complexes. The results indicated that the selectivity of lidorestat against AKR1B1 protein was mainly due to polar interactions with residues such as TRP20, TYR48, TRP111, PHE122, TRP219 and LEU300, while the selectivity mechanism of HAHE against AKR1B10 protein was based on polar interactions with TRP21, TYR49, HIS111, TRP112, PHE123, TRP220, ASN300 and VAL301 (Fig. 11). However, due to the limitations of binding affinity calculations in accurately determining selectivity attribution, it is plausible to hypothesize that the specificity of AKR1B inhibitors is influenced by the distinct structural conformations of the binding pocket in AKR1B1/10 and the variations in the hydrophobic region.

3.5. Alanine scanning mutagenesis analysis

Afterwards, to further gain rational insights into the key residues impact in the activation region of the AKR1B1-NADP$^+$-lidorestat

and AKR1B10-NADP$^+$-HAHE complexes, an alanine scanning mutagenesis analysis was conducted. The intermolecular interactions, including the AKR1B1 residues TRP20, TYR48, HIS110, TRP111, PHE122, TRP219, and LEU300, and the AKR1B10 residues TRP21, TYR49, HIS111, TRP112, PHE123, TRP220, ASN300 and VAL301 were selected to mutate to alanine (Fig. 12). Positive $\Delta\Delta G$ values indicate that wild-type residues interact more favorably with inhibitors than the mutated residues. Mutations weaken protein–inhibitor interactions in both complexes. Mutations like W111A, F122A and L300A in the AKR1B1/lidorestat complex and mutations like F123A and Y49A in the AKR1B10/HAHE complex did not affect their binding behaviour, suggesting that these residues primarily interact with ligands through backbone atoms rather than side chains. The alterations in binding free energy result from mutations in core residues W20A, Y48A, H110A and W219A in AKR1B1/lidorestat, as well as W21A, H111A, W112A, W220A, N300A and V301A in AKR1B10/HAHE, highlighting the significance of unmutated residues. It seems reasonable to conclude that mutations of the above residues into alanine shorten the side chain length of
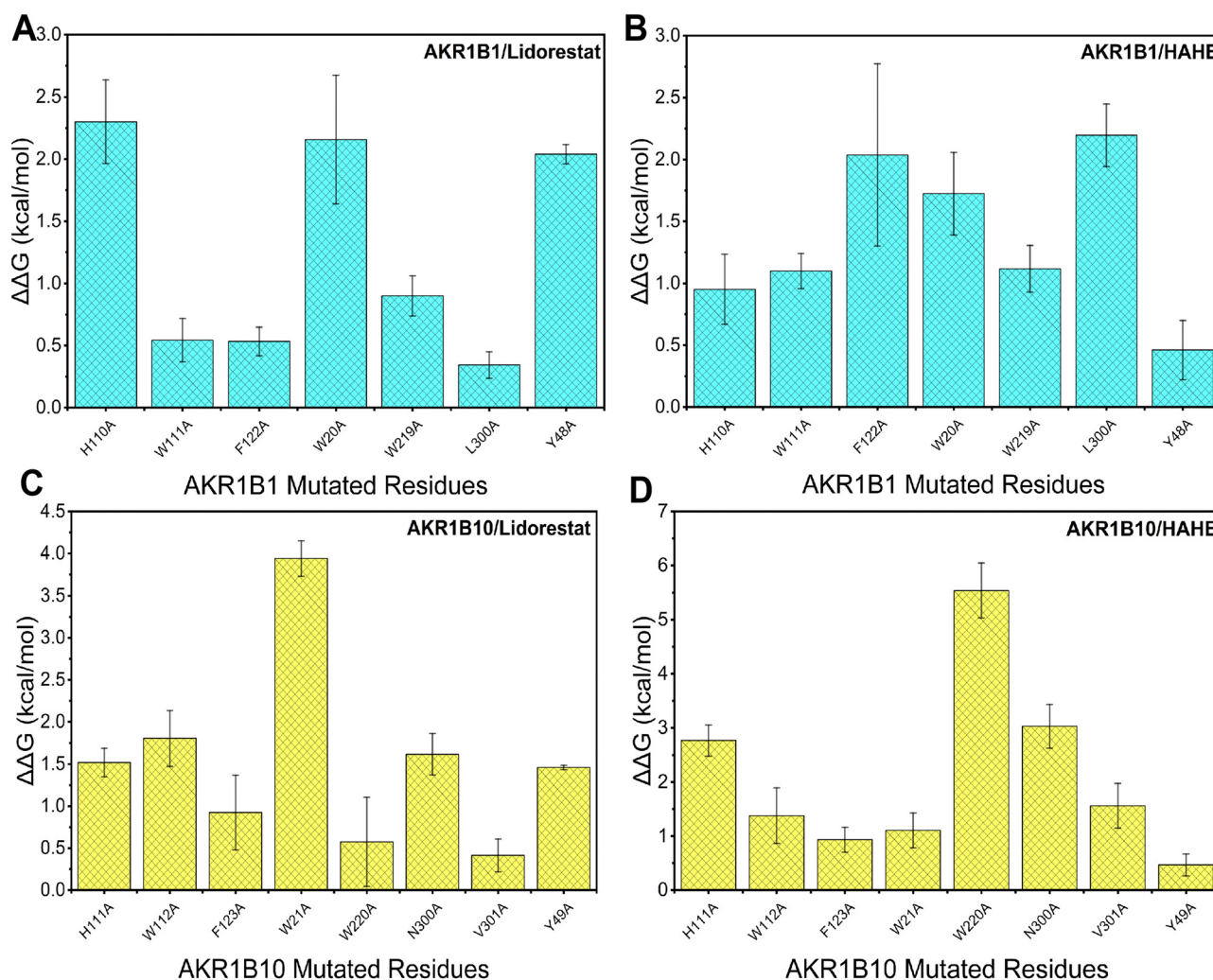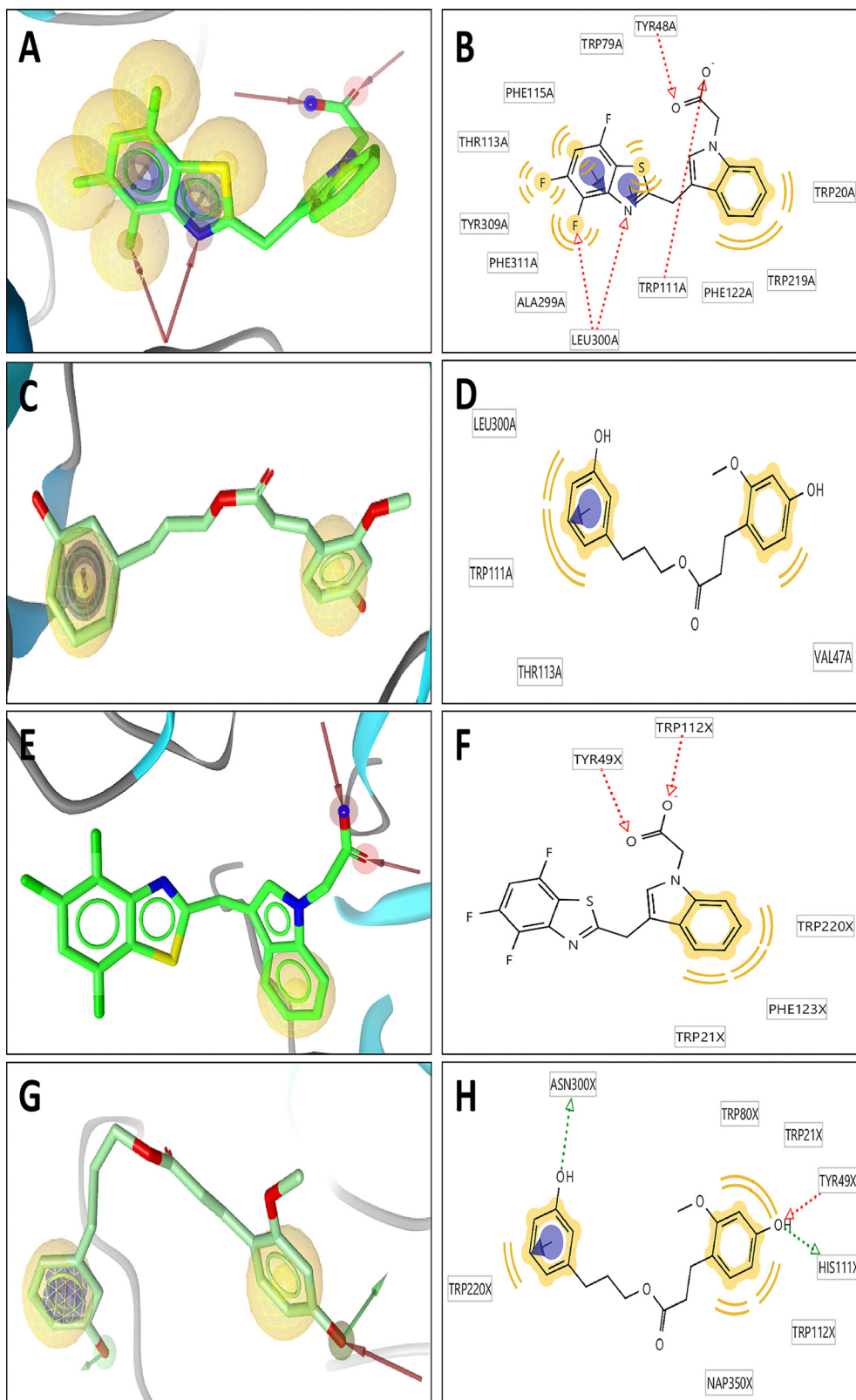

Fig. 12 Alanine scanning mutagenesis analysis of (A) AKR1B1/lidorestat, (B) AKR1B1/HAHE, (C) AKR1B10/lidorestat and (D) AKR1B10/HAHE complexes.

Fig. 13 Three-dimensional and two-dimensional pharmacophore modeling derived from the two X-ray structures of AKR1B1 and AKR1B10 in complex with lidorestat and HAHE, respectively. (A) and (B) AKR1B1/lidorestat. (C) and (D) AKR1B1/HAHE. (E) and (F) AKR1B10/lidorestat. (G) and (H) AKR1B10/HAHE. The pharmacophore features were represented in LigandScout by color codes in which, hydrogen bond acceptors, hydrogen bond donors, hydrophobic regions, ionizable positive charge and exclusion volume depicted as red scissor, green scissor, yellow spheres and blue spheres, respectively.

residues, thereby reducing interaction with ligands. Combined with per-residue energy decomposition analysis, it can be inferred that ligands binding with the AKR1B1 residues TRP111 and LEU300 tend to exhibit greater selectivity as AKR1B1 inhibitors, while ligands binding with the AKR1B10 residue TRP220 tend to display enhanced selectivity as AKR1B10 inhibitors.

3.6. Pharmacophore modeling of key chemical features

Pharmacophore is defined as the collection of spatial and electronic features, including hydrogen bond acceptors, hydrogen bond donors, hydrophobic regions, cationic sites, anionic sites, aromatic rings, and repulsors. It is necessary to ensure optimal supramolecular interaction with a particular biological target and trigger (or block) its biological response. To further explain the selectivity mechanisms of AKR1B1/10 inhibition, pharmacophore features were analyzed to better understand the requirements of amino acid residues within the AKR1B1/10 active site. As shown in Fig. 13, there were four hydrogen bond acceptors, six hydrophobic interaction areas and two aromatic rings in the AKR1B1/lidorestat complex, indicating that the carboxyl oxygen atom of lidorestat forms two hydrogen bonds with the catalytic subunits TYR48 and TRP111 of AKR1B1, and the nitrogen atom and fluorine atom on benzothiazole formed additional hydrogen bonds with LEU300. However, two hydrophobic interaction areas and an aromatic ring were observed in the AKR1B1/HAHE complex and did not form strong hydrogen bonds with protein, resulting in a poor affinity towards protein. For the AKR1B10 system, there were a hydrogen bond acceptor, two hydrogen bond donors, two hydrophobic interaction areas and an aromatic ring in the AKR1B10/HAHE complex. The hydroxyl oxygen atom in HAHE formed two hydrogen bonds with TYR49 and HIS111 of AKR1B10, and the hydroxyl oxygen atom in its benzene ring formed additional hydrogen bonds with ASN300. Nevertheless, two hydrogen bond acceptors and a hydrophobic interaction area were observed in the AKR1B10/lidorestat complex, indicating that the carboxyl oxygen atom of lidorestat formed two hydrogen bonds with TYR49 and HIS111 of AKR1B10, respectively. Together, these results further confirmed the results of molecular docking and MD simulation.

4. Conclusions

In the present study, we utilized a series of computational methods, such as sequence alignment, structural comparison, Protein Contacts Atlas analysis, molecular docking, MD simulation, MM-GBSA calculation, alanine scanning mutagenesis and pharmacophore model analysis to explore and validate the structural characteristics of crucial amino acid compositions between AKR1B1/10 and their representative selective inhibitor, as well as to elucidate the selectivity mechanisms of AKR1B1/10 inhibition. These crucial amino acids within the binding pocket, namely TRP20, TYR48, HIS110, TRP111, PHE122, TRP219 and LEU300 for AKR1B1 and TRP21, TYR49, HIS111, TRP112, TRP220, ASN300 and VAL301 for AKR1B10, exhibited

the ability to establish robust interactions with the selective inhibitors of AKR1B1/10, which significantly contributed to the selective mechanism of the inhibitors. The perspective offered in this research seeks to suggest a novel concept for AKR1B1/10 selective inhibitors while acknowledging the inherent constraints of computational chemistry techniques and the lack of supplementary chemical and biological tests to authenticate the presented arguments. In conclusion, these findings offer potential for elucidating the mechanism of AKR1B1/10 selection, thereby informing future approaches for the rational development of selective inhibitors targeting AKR1B1/10.

Author contributions

Conceptualization, M. L. and Z. W.; methodology, M. L., X. Q., J. L., Y. J., J. J., J. G.; software, M. L., X. Q. and J. L.; validation, H. X., Y. W. and H. B.; formal analysis, M. L. and X. Q.; investigation, M. L. and J. L.; resources, Z. W.; data curation, Z. W.; writing – original draft preparation, M. L.; writing – review and editing, M. L., X. Q. and Z. W.; visualization, Z. W.; supervision, Z. W.; project administration, M. L.; funding acquisition, M. L. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 N. P. Syamprasad, S. Jain, B. Rajdev, N. Prasad, R. Kallipalli and V. G. M. Naidu, *Biochem. Pharmacol.*, 2023, **211**, 115528.

2 T. M. Penning, S. Jonnalagadda, P. C. Trippier and T. L. Rizner, *Pharmacol. Rev.*, 2021, **73**, 1150–1171.

3 M. Singh, A. Kapoor and A. Bhatnagar, *Metabolites*, 2021, **11**, 655.

4 T. M. Penning, *Chem. – Biol. Interact.*, 2015, **234**, 236–246.

5 S. Thakur, S. K. Gupta, V. Ali, P. Singh and M. Verma, *Arch. Pharm. Res.*, 2021, **44**, 655–667.

6 Y. Wang, J. Fan, Y. Tong, L. Wang, L. Wang, C. Weng, C. Lai, J. Song and W. Zhang, *Comput. Biol. Med.*, 2023, **158**, 106740.

7 C. Turkes, M. Arslan, Y. Demir, L. Cocaj, A. R. Nixha and S. Beydemir, *J. Mol. Recognit.*, 2022, **35**, e2991.

8 E. Pastel, J. C. Pointud, A. Martinez and A. M. Lefrancois-Martinez, *Front. Endocrinol.*, 2016, **7**, 97.

9 R. Khayami, S. R. Hashemi and M. A. Kerachian, *J. Cell. Mol. Med.*, 2020, **24**, 8890–8902.

10 O. Gallego, F. X. Ruiz, A. Ardevol, M. Dominguez, R. Alvarez, A. R. de Lera, C. Rovira, J. Farres, I. Fita and X. Pares, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20764–20769.

11 S. Banerjee, *Adv. Exp. Med. Biol.*, 2021, **1347**, 65–82.

12 M. Hojnik, N. K. Suster, S. Smrkolj, D. Sisinger, S. F. Grazio, I. Verdenik and T. L. Rizner, *Cancers*, 2022, **14**, 809.

13 K. R. Zhang, Y. F. Zhang, H. M. Lei, Y. B. Tang, C. S. Ma, Q. M. Lv, S. Y. Wang, L. M. Lu, Y. Shen, H. Z. Chen and L. Zhu, *Sci. Transl. Med.*, 2021, **13**, eabg6428.

14 Y. Su, P. Song, H. Wang, B. Hu, J. Wang and M. S. Cheng, *J. Biomol. Struct. Dyn.*, 2020, **38**, 3825–3837.

15 M. Foppiano and G. Lombardo, *Lancet*, 1997, **349**, 399–400.

16 N. Hotta, R. Kawamori, M. Fukuda, Y. Shigeta and G. Aldose, Reductase Inhibitor-Diabetes Complications Trial Study, *Diabetic Med.*, 2012, **29**, 1529–1533.

17 C. Bailly, *Eur. J. Pharmacol.*, 2022, **931**, 175191.

18 M. C. Van Zandt, M. L. Jones, D. E. Gunn, L. S. Geraci, J. H. Jones, D. R. Sawicki, J. Sredy, J. L. Jacot, A. T. Dicioccio, T. Petrova, A. Mitschler and A. D. Podjarny, *J. Med. Chem.*, 2005, **48**, 3141–3152.

19 M. Soda, D. Hu, S. Endo, M. Takemura, J. Li, R. Wada, S. Ifuku, H. T. Zhao, O. El-Kabbani, S. Ohta, K. Yamamura, N. Toyooka, A. Hara and T. Matsunaga, *Eur. J. Med. Chem.*, 2012, **48**, 321–329.

20 G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe, Jr., *Pharmacol. Rev.*, 2014, **66**, 334–395.

21 E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang and T. Hou, *Chem. Rev.*, 2019, **119**, 9478–9508.

22 I. C. Simoes, I. P. Costa, J. T. Coimbra, M. J. Ramos and P. A. Fernandes, *J. Chem. Inf. Model.*, 2017, **57**, 60–72.

23 C. UniProt, *Nucleic Acids Res.*, 2019, **47**, D506–D515.

24 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.

25 M. Kayikci, A. J. Venkatakrishnan, J. Scott-Brown, C. N. J. Ravarani, T. Flock and M. M. Babu, *Nat. Struct. Mol. Biol.*, 2018, **25**, 185–194.

26 Maestro version 13.5, Schrödinger, LLC, New York, NY, 2023.

27 G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput. Aided Mol. Des.*, 2013, **27**, 221–234.

28 M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E. Shaw and R. A. Friesner, *Proteins*, 2004, **55**, 351–367.

29 C. Lu, C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G. A. Ross, M. K. Dahlgren, E. Russell, C. D. Von Bargen, R. Abel, R. A. Friesner and E. D. Harder, *J. Chem. Theory Comput.*, 2021, **17**, 4291–4300.

30 LigPrep, Schrödinger, LLC, New York, NY, 2023.

31 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.

32 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J. Med. Chem.*, 2004, **47**, 1750–1759.

33 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2006, **49**, 6177–6196.

34 K. J. Bowers, D. E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan and D. E. Shaw, Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters, *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 2006, pp. 43–43.

35 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **3**, 198–210.

36 J. Luan, B. Hu, S. Wang, H. Liu, S. Lu, W. Li, X. Sun, J. Shi and J. Wang, *Phys. Chem. Chem. Phys.*, 2022, **24**, 17105–17115.

37 I. Massova and P. A. Kollman, *J. Am. Chem. Soc.*, 1999, **121**, 8133–8143.

38 G. Wolber and T. Langer, *J. Chem. Inf. Model.*, 2005, **45**, 160–169.

39 L. Zhang, H. Zhang, Y. Zhao, Z. Li, S. Chen, J. Zhai, Y. Chen, W. Xie, Z. Wang, Q. Li, X. Zheng and X. Hu, *FEBS Lett.*, 2013, **587**, 3681–3686.