frontiers | Frontiers in Drug Discovery

TYPE Brief Research Report
PUBLISHED 28 January 2026
DOI 10.3389/fddsv.2026.1770904

OPEN ACCESS

EDITED BY
José L. Medina-Franco,
National Autonomous University of Mexico,
Mexico

REVIEWED BY
Felipe Victoria-Muñoz,
Fundación Universitaria Salesiana, Colombia
Alberto Marban,
National Autonomous University of Mexico,
Mexico

*CORRESPONDENCE
Pedro J. Ballester,
p.ballester@imperial.ac.uk

RECEIVED 18 December 2025
REVISED 12 January 2026
ACCEPTED 13 January 2026
PUBLISHED 28 January 2026

CITATION
James N and Ballester PJ (2026) Structure-based virtual screening for TRPM8 modulators. *Front. Drug Discov.* 6:1770904.
doi: 10.3389/fddsv.2026.1770904

Structure-based virtual screening for TRPM8 modulators

Nivya James and Pedro J. Ballester*

Department of Bioengineering, Imperial College London, London, United Kingdom

Transient receptor potential melastatin 8 (TRPM8) is an emerging therapeutic target, yet the performance of available structural models for docking and optimal docking protocols for virtual screening (VS) remains unclear. Here, we benchmarked two available TRPM8 structural conformations (agonist-bound TRPM8^{7WRE} and antagonist-bound TRPM8^{9B6G}) using docking tools, Smina and rDock, using known TRPM8 inhibitors and property-matched decoys. rDock achieved the highest hit rates and outperformed Smina in ranking true actives at first-ranks than their corresponding decoys, whereas Smina showed a target-structure dependence on docking performance but delivered superior overall ranking quality across both target structures. Both docking tools displayed considerable overlap between active and decoy score distributions, indicating only moderate discriminatory power of docking scores alone. When prioritizing a small subset of top-ranked compounds, integrated screening approaches, particularly the consensus protocol, improved the recovery of true actives, while the hierarchical protocol achieved comparable performance at a substantially lower computational cost. Collectively, this work establishes a reproducible VS benchmark for TRPM8 and supports the use of different screening protocols to improve early hit identification.

KEYWORDS

benchmarking, ion channels, molecular docking, TRPM8, virtual screening

Introduction

In the early 2000's, studies aimed at understanding the molecular basis of cold sensation revealed TRPM8 as a key cold-activated ion channel. This ground-breaking discovery was awarded the 2021 Nobel Prize for Physiology and Medicine to David Julius and Ardem Patapoutian. TRPM8 is a polymodal, nonselective cation channel activated by cold temperatures and cooling agents such as menthol, and it can be modulated by both agonists and antagonists. Since its discovery, TRPM8 has become one of the most extensively studied TRP channels (Wu et al., 2024). Subsequent research has expanded its functional repertoire beyond thermo-sensation to include roles in cold-induced and neuropathic pain Weyer and Lehto, 2017; Bianchini et al., 2021) multiple cancers (Ochoa et al., 2023), inflammation (Ramachandran et al., 2013) and a range of additional pathological conditions (Liu et al., 2020).

As of 28 May 2025, the U.S. Food and Drug Administration (FDA) approved the first-in-class drug targeting TRPM8 for the treatment of dry-eye disease (Zhou et al., 2025), nearly two decades after the channel's initial discovery. This milestone underscores the need for more cost-effective and efficient strategies, such as VS, to accelerate the identification of early hit molecules for TRPM8. However, there is a notable lack of published studies evaluating optimal VS methodologies for this target.

Building on this need for more efficient drug lead discovery, the present study focuses on defining best practices for structure-based VS against TRPM8. Specifically, we aim to

TABLE 1 Composition of the TRPM8 benchmark dataset.

Data source	Total downloaded	After data cleaning	Actives	Agonists	Antagonists
PubChem	1,315	565	564	125	439
ChEMBL	1,168	747	533	72	461
Binding database	2,645	2,291	1725	113	1,612
Final merged set	2,553		1984	135	1849

Molecules retrieved from PubChem, ChEMBL, and BindingDB underwent an initial curation step in which entries lacking SMILES strings, activity values, assay descriptions, or with relationship types other than " = " or ">" were removed. The three curated files were now merged, and duplicates (identical SMILES and activity values) were removed and one ID per molecule was retained. Prior to merging and de-duplication, SMILES strings were standardized using RDKit to ensure consistent molecular representations across data sources. This resulted in a final curated set of 2,553 unique molecules, comprising of 1,984 actives. For benchmarking, average of 48 decoys per active were generated using DeepCoy method, yielding the final VS dataset of 96,634 molecules (active-to-Decoy ratio is hence 1:48).

determine which available TRPM8 structural model(s) offer the most reliable basis for such VS, which molecular docking protocol(s) yield the highest predictive performance, and how different docking methods can be optimally combined to enhance hit identification. By systematically addressing these questions, this work also introduces a robust and reproducible VS benchmark tailored to TRPM8.

Methods

Preparing virtual screening benchmark for TRPM8

To build this TRPM8 virtual screening (VS) benchmark, we compiled known TRPM8 modulators from three publicly available databases: PubChem, ChEMBL, and BindingDB (access date: June 2024; Table 1). All retrieved molecules were processed using a previously established data-curation workflow (Tran-Nguyen et al., 2023). Reported activity measurements were converted to their corresponding negative logarithmic values (p*Act*; Equation 1), and the median pAct value for each compound was used as its final activity metric. Compounds with p*Act* ≥ 6 (corresponding to activity ≤1 μM) were classified as actives. Only compounds classified as actives were retained for benchmarking. We assume that the reported TRPM8 activity values reflect ligand interactions at the well-characterized ligand-binding pocket within the voltage-sensing–like domain (VSLD), which mediates the interaction of most known TRPM8 modulators (Xu et al., 2020). This assumption is supported by available cryo-electron microscopy structures complexed with either the agonist icilin or the antagonist AMTB.

$$pAct = -\log_{10}(IC_{50} \text{ or } K_d \text{ or } EC_{50}) \quad (1)$$

Agonist and antagonist annotations were assigned using a combination of functional assay descriptions and activity type (IC_{50}, K_d, and EC_{50}) information, depending on data availability. For compounds retrieved from ChEMBL and PubChem, functional assay descriptions were manually inspected to ensure that reported EC_{50} and IC_{50} values corresponded to receptor activation or inhibition, respectively. Compounds with ambiguous assay descriptions or unclear activity types (e.g., Kb or unspecified activity annotations) were excluded. For compounds retrieved from BindingDB, where functional assay descriptions were not available, classification was based on the reported activity types and their potencies. EC_{50} values were used to identify agonists, while IC_{50} values were used to identify antagonists. For compounds with multiple reported activity types, pEC_{50} and pIC_{50} values were compared, and compounds were classified according to the activity type associated with the stronger potency. For compounds with both pK_d and pEC_{50} values, compounds with pK_d exceeding pEC_{50} were classified as antagonists. The curated dataset was dominated by IC_{50} measurements (~2,300 compounds), a smaller number of EC_{50} measurements (~220 compounds) and two K_d values, with only minimal overlap between assay types.

We gathered and curated a total of 1,984 TRPM8-active compounds. To generate a challenging class-imbalanced dataset, we aimed at generating 50 property-matched decoys per active using DeepCoy (Imrie et al., 2021), a graph-based deep-learning approach that designs tailored decoys with similar physicochemical profiles but distinct topological structures. The resulting chemical space and property distributions are shown in Supplementary Figures S1A,B. As usual, not all the requested decoys were successfully generated. This resulted in an average of 48 decoys per active, leading to a final benchmark dataset comprising 1,984 TRPM8 actives and 94,650 decoys.

Selection of TRPM8 structures

A total of 26 TRPM8 structures were available in the PDB (June 2024), all solved by cryo–electron microscopy. Of these, only one corresponds to the human protein (apo), while the remaining 25 are of mice (18) and avian (7) origin. For structure-based docking, we focused exclusively on ligand-bound target structures. Because mouse TRPM8 provides multiple ligand-bound structures and shares ~94% sequence identity with the human TRPM8 channel (Yin et al., 2022), these structures were selected for docking studies. Among these, only those with the highest resolution (<3 Å) were retained.

Given that TRPM8 mediates both agonist and antagonist activities, we incorporated structural diversity in our docking experiments by using two distinct TRPM8 conformations: an agonist-bound structure $TRPM8^{7WRE}$ (PDB ID: 7WRE; icilin-bound; 2.5Å) and an antagonist-bound structure $TRPM8^{9B6G}$ (PDB ID: 9B6G; AMTB-bound; 2.81Å). This dual-structure strategy enables us to capture the conformational variability of the TRPM8 binding site and to evaluate which structural state provides the most reliable basis for VS.

Docking preparation

We prepared the selected TRPM8 structures using the DockPrep module in UCSF Chimera v1.17.3. All non-standard residues were removed from the TRPM8 structures. The TRPM8^{7WRE} structure contains a Ca^{2+} ion located near the ligand-binding site within the calcium-binding pocket (Zhao et al., 2022). This ion was retained to preserve the agonist-bound conformation. Further, solvent molecules were removed, hydrogens were added and optimized for hydrogen bonding. Partial charges were assigned using the Gasteiger method, and any missing side chains were rebuilt using the Dunbrack 2010 rotamer library (Shapovalov and Dunbrack, 2011).

Each of the 96,634 molecules was prepared using the RDKit cheminformatics toolkit (version 2024.09.1). Salts were removed using the SaltRemover module, and hydrogens were added to the resulting chemical structures. One single 3D conformer per molecule was then generated using the ETKDGv3 method (Riniker and Landrum, 2015) employing the EmbedMolecule function. A fixed random seed (0xf00d) was applied to ensure reproducibility and the generated conformers were written in SD format.

rDock protocol

rDock (Ruiz-Carmona et al., 2014) is an open-source docking program originally developed for VS against RNA targets but later extended to support protein–ligand docking. Its sampling protocol combines a genetic algorithm with Monte Carlo perturbations followed by local optimization. rDock employs an empirical scoring function (SF) with molecular-mechanics terms estimating the van der Waals, electrostatic, hydrogen-bonding and hydrophobic contributions to binding, with an optional desolvation term. The default scoring function (SF3) excludes the desolvation term and incorporates a repulsive polar penalty and works better with protein targets (Ruiz-Carmona et al., 2014). Accordingly, all docking scores in this work were computed using the default SF3 scoring function.

To analyze the molecules with rDock, we generated first an active-site parameter file specifying the binding site, search space, receptor flexibility, and SF. This file was processed using the rbcavity tool to build the receptor grid around the binding pocket. rDock also allows limited receptor flexibility during parameter-file generation, enabling rotation of terminal–OH and–NH_3^+ groups on residues within a user-defined distance from the binding site.

To determine the optimal docking settings, redocking was performed for each protein structure, and the parameter configuration yielding the lowest RMSD between the bound and docked ligand were selected. Several parameters and combinations for docking box (6 Å, 10 Å and 15 Å) and binding site flexibility (3 Å, 5 Å and 6 Å) were tested. In the final parameter files, residues within 5 Å of the binding site were treated as flexible for TRPM8^{7WRE}, whereas a 3 Å flexibility radius was used for TRPM8^{9B6G}. For both structures, the docking cavity was defined as a 10 Å region surrounding the bound ligand.

Smina protocol

We also dock the same molecules using Smina (Koes et al., 2013), a fork of AutoDock Vina (Trott and Olson, 2010). Smina employs a SF derived from AutoDock Vina, which is an empirical scoring model incorporating Gaussian steric terms, hydrophobic interactions, hydrogen bonding, and a rotatable-bond penalty. Ligand sampling is carried out using an iterated local search algorithm that combines stochastic perturbations with gradient-based local optimization to identify low-energy binding poses within the protein pocket. The degree of conformational search was controlled by the exhaustiveness parameter, where higher values perform more extensive sampling at the cost of increased computational time (Agarwal and Smith, 2023).

Smina docking parameters were optimized through redocking experiments analogous to those performed with rDock. Several values of exhaustiveness (1, 8, and 16) and search box dimensions (10 Å, 15 Å, 30 Å, and 35 Å) were evaluated (data not shown). Based on these tests, an exhaustiveness value of 8 was selected for TRPM8^{7WRE} whereas an exhaustiveness value of 1 was selected for TRPM8^{9B6G} combined with a 15 Å × 15 Å × 15 Å search box for both the protein structures, as these settings provided the best balance between pose recovery and computational efficiency. These parameters were therefore used for all subsequent large-scale VS calculations. Prior to docking the dataset molecules, they were converted to MOL2 format from SD format using OpenBabel v3.0.0. For each ligand, Smina was configured to generate and output a single best-scoring pose, selected from the full conformational search, and the predicted binding affinity of this pose was used for downstream performance evaluation.

Consensus protocol

Combining results from different docking programs to generate a consensus score or rank has been reported to improve VS performance (Chang et al., 2010; Charifson et al., 1999; Scardino et al., 2025). In this study, we used each docking tool to rank the benchmark molecules by decreasing predicted potency (i.e., lower ranks are assigned to molecules predicted to have more potent TRPM8 activity). To form the rDock-Smina consensus model, for each molecule, we averaged its two ranks, one from rDock and the other from Smina. The performance of this consensus protocol was then evaluated by examining the enrichment within the top 1% of the ranked list.

Hierarchical protocol

We also implemented a hierarchical protocol, in which molecules were filtered sequentially using the two docking tools. This approach is motivated by the observation that rDock is computationally faster, whereas Smina prioritizes more potent molecules (Supplementary Figure S2). First from the rDock run, we selected the top 10% of molecules using their rDock scores. We then retrieved the Smina scores of this rDock-filtered subset and further selected the top 10% using their Smina scores. This results in a subset with the top 1% of the molecules by this protocol, on which

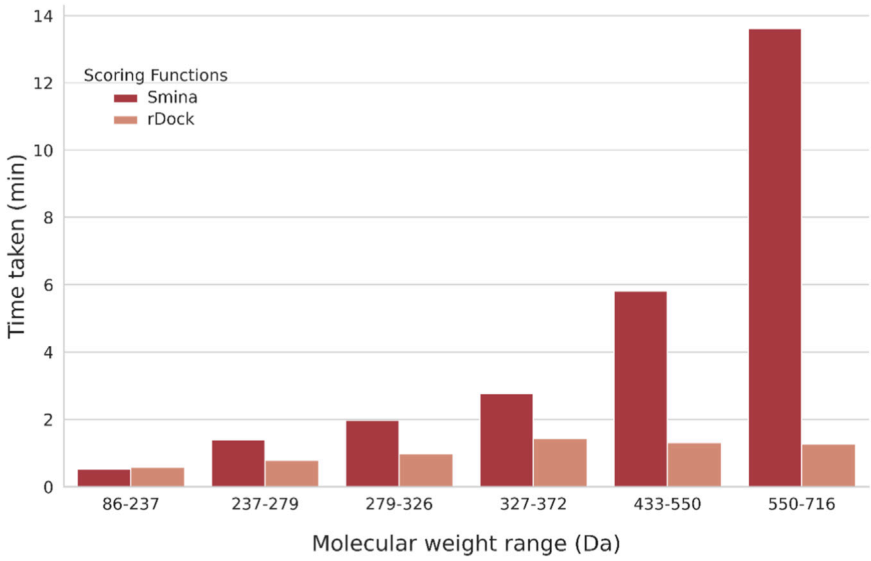

FIGURE 1
Docking runtimes of Smina and rDock across six molecular weight–based ligand subsets for both protein targets. Bars represent the average total docking time per molecular weight subset (averaged over the two targets) under a single-core, single-CPU setup.

performance was evaluated. This protocol is practically as fast as rDock (the overhead is the time to docking 10% of the molecules with Smina).

Virtual screening performance metrics

The primary metric used for evaluating VS performance was the hit rate within the top 1% of the ranked molecules. Hit rate is defined as the proportion of true active compounds retrieved among the top-ranked candidates generated by the docking protocols. A higher hit rate indicates greater efficiency of a given protocol in identifying active hits (Equation 2). We also calculate this metric at the top 0.5%.

$$\text{Hit Rate (\%)} = \frac{Number\ of\ true\ actives}{Total\ number\ of\ compounds\ screened} \times 100 \quad (2)$$

In addition to hit rate, we analyzed the predicted binding affinity scores across the full dataset to assess each protocol's ability to distinguish true positives from false positives. The time to screen the molecules was taken as the CPU time required to complete that docking run.

Results

rDock is up to 4x times faster than Smina

The docking speeds of Smina and rDock were evaluated using a single-core, single-CPU setup. The ligand dataset was divided into six subsets according to molecular weight (MW), with each subset representing a distinct MW range. rDock consistently outperformed Smina in docking speed. As illustrated in Figure 1, rDock maintained near-constant docking times across all MW ranges, whereas Smina exhibited a strong dependence on ligand size, with docking time increasing progressively with MW.

Out of 96,634 ligands submitted for docking, rDock successfully docked 96,614 dataset molecules against both TRPM8^{7WRE} and TRPM8^{9B6G}. This corresponds to a docking success rate of 99.98% for each target, defined as the fraction of ligands for which a valid docking pose and score were generated. Whereas Smina docked 96,634 ligands against TRPM8^{7WRE} and 95,453 ligands against TRPM8^{9B6G}, corresponding to success rates of 100% and 98.78%, respectively. These results indicate good computational stability of both docking tools for large-scale VS. In terms of computational efficiency, rDock completed the full docking process in approximately 1 h, whereas Smina required approximately 4 h, making rDock about four times faster than Smina on this dataset.

rDock also outperforms Smina in ranking actives higher than their decoys

rDock outperformed Smina by ranking a larger number of active molecules at the top position among their respective decoy sets (TRPM8^{7WRE}: rDock- 442; Smina- 317; and TRPM8^{9B6G}: rDock- 442; Smina- 294), reflecting superior early enrichment against both target structures (Figure 2). To further summarize overall ranking performance, we define here the target-wise weighted average rank (WAR) for both Smina and rDock using Equation 3, where a lower value indicates better overall ranking quality.

$$\text{WAR} = \frac{\Sigma\,(rank \times number\ of\ active\ at\ that\ rank)}{Number\ of\ actives} \quad (3)$$

For TRPM8^{7WRE}, Smina and rDock achieved WARs of 12.91 and 16.86, respectively, while for TRPM8^{9B6G}, the corresponding WARs

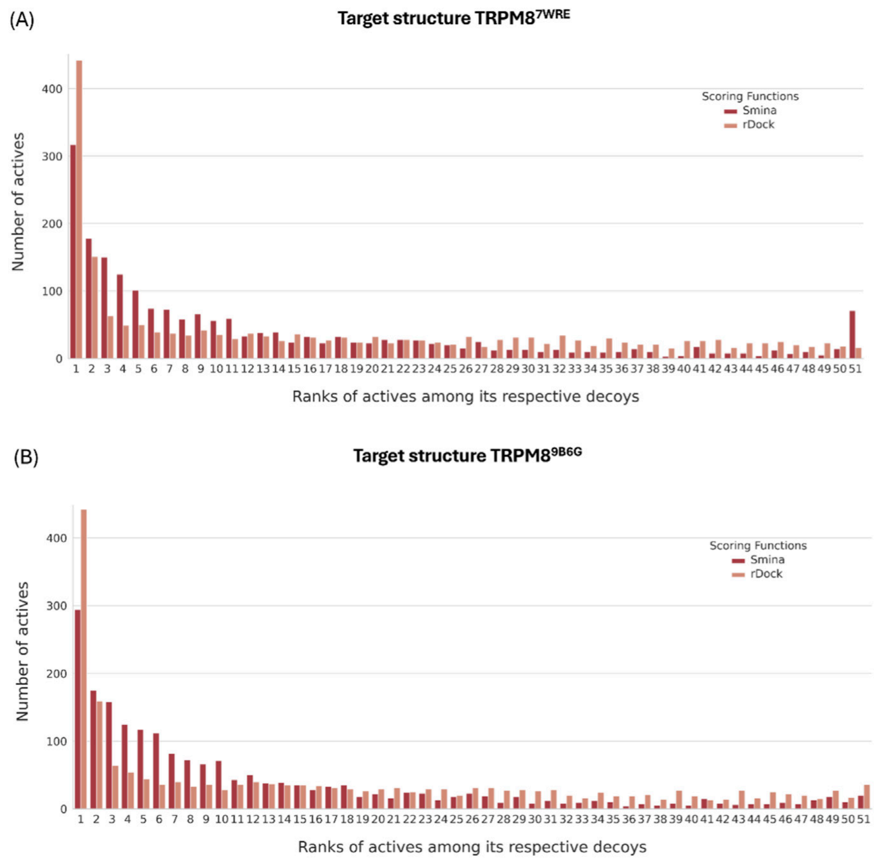

FIGURE 2
Ranking performance of Smina and rDock on two TRPM8 target structures. Grouped bar charts illustrate the number of active molecules observed at each rank as predicted by rDock and Smina, when docked to **(A)** TRPM8^{7WRE} and **(B)** TRPM8^{9B6G} structure. Indicate the number of actives that were scored better than all their respective decoys, whereas the highest ranks (rightmost, up to rank 51) indicate actives that were scored worse than their respective decoys. The y-axis reports the number of actives observed at each rank.

were 11.62 and 16.61. These results indicate that, despite rDock exhibiting stronger first-rank enrichment, Smina achieves superior overall ranking quality across both targets. This is consistent with Figure 2, where Smina consistently ranks a large fraction of actives within approximately the top 2–15 positions relative to their associated decoys.

The hierarchical protocol provides the best efficiency-hit rate compromise

rDock demonstrated similar hit rates across both TRPM8 conformations, whereas Smina performed better against TRPM8^{9B6G} than against TRPM8^{7WRE}. At a hit rate of 1%, rDock achieved the highest performance across all protocols and structures, while Smina exhibited the lowest (Supplementary Figures S3A,B). At a hit rate of 0.5% (Figure 3), the consensus and hierarchical protocols showed the highest performance against TRPM8^{9B6G} (38% and 28%, respectively; Figure 3B), whereas the consensus protocol performed best against TRPM8^{7WRE} (30%, Figure 3A). These results indicate the benefit of using the integrated protocols to improve prioritization of true actives when selecting a small, high-priority subset of compounds for downstream experimental testing, compared to individual docking tools.

Analysis of the predicted docking score distributions revealed that rDock produced extreme range of scores, that is more extreme negative values, than Smina for TRPM8 structures (Supplementary Figures S3C,D). However, for both docking tools, the score distributions of true positives and false positives overlapped substantially, with very similar median values. This indicates that neither scoring function was able to reliably separate active molecules from decoys based solely on predicted docking scores, among the top 1% ranked compounds.

To further assess the added value of hierarchical and consensus protocols, we examined the overlap of true actives between methods at the top 0.5% of the ranked molecules for each method using UpSet plots (Figures 3C,D). rDock identified the largest number of unique or non-redundant active molecules (TRPM8^{7WRE}: 80, TRPM8^{9B6G}: 73), followed by consensus protocol (TRPM8^{7WRE}: 61, TRPM8^{9B6G}: 50) for both target structures, while hierarchical protocol showed considerable overlap with consensus protocol (TRPM8^{7WRE}: 41, TRPM8^{9B6G}: 82). Although Smina identified a small number of unique actives (TRPM8^{7WRE}: 14, TRPM8^{9B6G}: 49), when considering rDock and Smina alone, it added few actives beyond

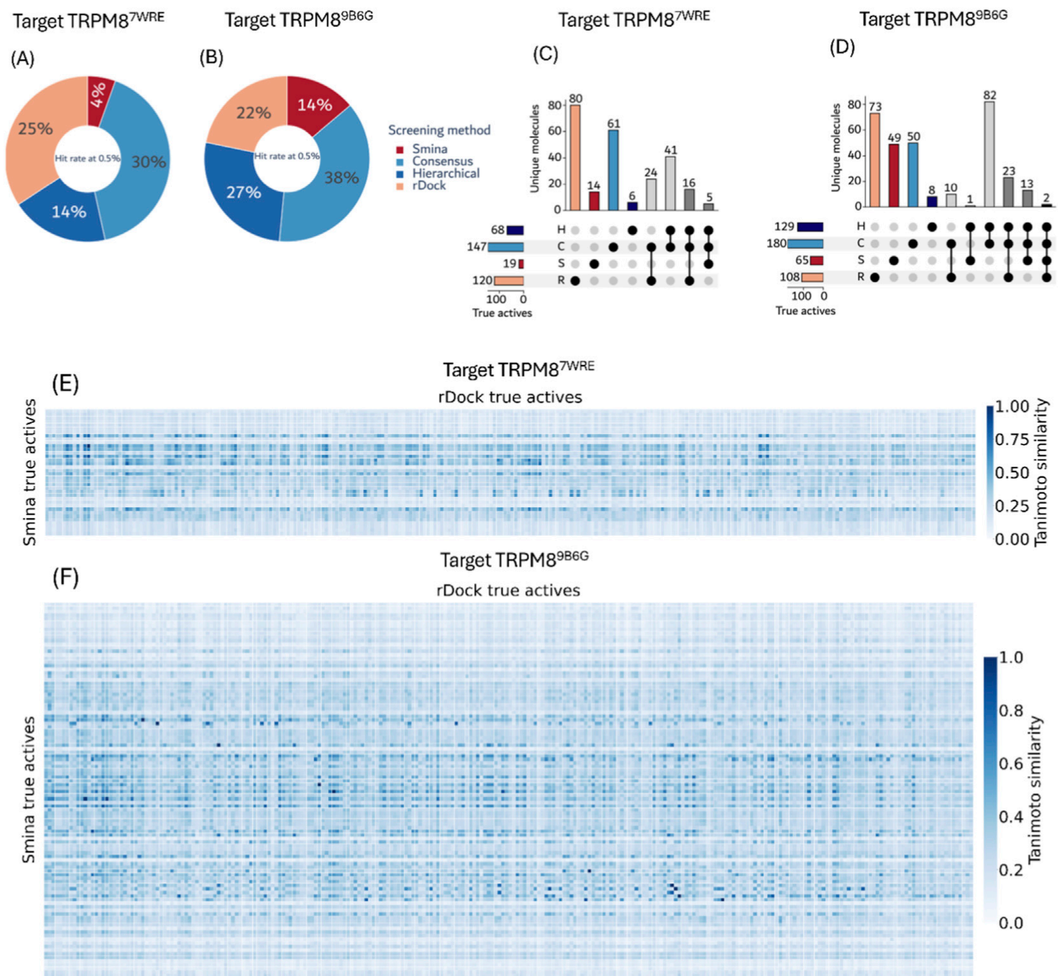

FIGURE 3
Comparison of VS performance and overlap analysis for Smina, rDock, Consensus protocol and Hierarchical protocol for both TRPM8 targets. The Donut plots show the hit rates at 0.5% for different screening methods against **(A)** TRPM8^{7WRE} and **(B)** TRPM8^{9B6G}. Donut plots showing the hit rates at 0.5% for different screening methods against **(A)** TRPM8^{7WRE} and **(B)** TRPM8^{9B6G}. Hit rates represent the fraction of known active compounds recovered within the top-ranked 0.5% of the screened library for each screening method, calculated independently. The values shown therefore do not sum to 100%, and for each method the remaining fraction corresponds to active compounds ranked outside the 0.5% cutoff. **(C,D)** UpSet plots illustrate the overlap of true actives identified by the four screening methods at a 0.5% cutoff for **(C)** TRPM8^{7WRE} and **(D)** TRPM8^{9B6G}. Bars represent the number of unique molecules retrieved by each individual method and their intersections, with the total number of true actives retrieved by each method shown on the left. **(E,F)** Pairwise Morgan fingerprint (ECFP4)-based Tanimoto similarity between Smina and rDock true actives for TRPM8^{7WRE} **(E)** and TRPM8^{9B6G} **(F)**, showing predominantly low-to-moderate similarity.

those already supported by either consensus or hierarchical protocols. This is reflected by the absence of an R–S only intersection in the UpSet plots. In contrast, both consensus and hierarchical screening recovered actives that were not identified by rDock and Smina alone. While consensus protocol therefore provides greater non-redundancy, it requires docking of the full compound library. We also provide UpSet plots for the less stringent 1% cutoff (Supplementary Figure S4).

From a practical screening perspective, these performances must be weighed against computational cost. Selecting the top 0.5% or 1% of compounds using rDock alone would require experimental testing of 483 or 966 molecules, respectively. However, hierarchical screening selects the same number of compounds as rDock at 0.5% and 1%, but includes additional true actives that are not prioritized by rDock alone. Importantly, this gain is achieved with limited additional computational expense, as the hierarchical protocol requires Smina docking of only 10% of the library (9,663 compounds), compared to full library docking for consensus protocol, thus providing an efficient compromise between enrichment performance and runtime.

Finally, Morgan fingerprint (ECFP4)-based pairwise Tanimoto similarity analysis between the true actives identified by Smina and rDock (Figures 3E,F) shows predominantly low-to-moderate similarity. This indicates that the two docking programs tend to retrieve chemically distinct compounds, consistent with limited scaffold overlap and the exploration of different regions of chemical space, as reported in previous studies (Guo et al., 2024). This suggests that compounds prioritized by the hierarchical protocol, which are supported by agreement between both docking tools, are likely to represent higher-confidence candidates for experimental validation.

Correctly predicted TRPM8 modulators are mostly antagonist

Despite the use of both agonist-bound and the antagonist-bound TRPM8 structure, neither docking protocol retrieved any known agonists. Using TRPM8^{7WRE}, Smina retrieved 36 true actives in the top-ranked subset, all of which were antagonists. Similarly, Smina docking against TRPM8^{9B6G} retrieved 104 true actives, again consisting exclusively of antagonists.

Consistent results were obtained with rDock, which retrieved 266 and 258 true actives against TRPM8^{7WRE} and TRPM8^{9B6G}, respectively, all of which were antagonists. Notably, none of the 135 compounds classified as agonists in the dataset were recovered among the top-ranked compounds by either docking protocol or either receptor structure.

Discussion

This study evaluated the suitability of two TRPM8 structural conformations (agonist-bound: TRPM8^{7WRE} and antagonist-bound: TRPM8^{9B6G}) and two docking protocols, Smina and rDock, for VS. Across the two TRPM8 structures, rDock showed consistent hit rates, whereas Smina performed better against TRPM8^{9B6G} than against TRPM8^{7WRE}, indicating that TRPM8^{9B6G} is the more suitable template for antagonist-focused screening with Smina.

In addition to performance differences, the two docking tools displayed distinct computational characteristics. rDock showed superior computational efficiency, with faster docking and minimal dependence on ligand molecular weight, whereas Smina docking time increased substantially with ligand size. These differences have practical implications for large-scale screening, particularly when computational resources are limited. At a 1% hit rate, rDock provided the strongest overall performance across targets, indicating its suitability for broad, cost-effective prioritization. However, at the more stringent 0.5% cutoff (representative of scenarios where only a very small number of compounds can be experimentally tested) consensus and hierarchical protocols consistently outperformed individual docking tools, demonstrating improved early enrichment. Notably, although rDock demonstrated stronger early recognition by placing more actives at first-rank, Smina achieved better overall ranking quality across targets. However, both docking tools showed considerable overlap between the score distributions of actives and decoys, indicating limited ability to reliably distinguish true actives solely based on docking scores.

The observation that rDock and Smina sample distinct regions of chemical space further supports the use of consensus and hierarchical protocols, as agreement between these diverse methods provides additional confidence in compound prioritization. Although consensus protocol retrieves more actives at 0.5% cutoff, this strategy is computationally expensive. In contrast, hierarchical screening improves the likelihood of selecting true actives without substantially increasing computational cost, thereby offering a practical balance between screening performance and efficiency.

Despite the use of both an agonist-bound and an antagonist-bound TRPM8 structure, and a ligand dataset containing compounds with both functional annotations, docking consistently retrieved only compounds that are experimentally classified as antagonists when functional annotations were examined *post hoc*. Notably, even when docking was performed against the agonist-bound TRPM8 structure and limited receptor flexibility was introduced in rDock, none of the 135 known TRPM8 agonists present in the dataset were recovered among the top-ranked compounds. This behavior likely reflects the strong class imbalance in the dataset (1,849 antagonists vs. 135 agonists), as well as inherent limitations of standard docking approaches in capturing the molecular features required for TRPM8 agonism. These observations highlight that docking scores primarily reflect binding compatibility rather than functional outcome.

Future work will therefore focus on developing target-specific SFs, which are expected to provide improved accuracy and functional prediction (Tran-Nguyen et al., 2023). In addition, predictive modelling strategies will be explored to anticipate which predicted binders will be agonist and which will be antagonist. Complementary analyses, such as the evaluation of ligand-based features, residue-level interaction patterns, and molecular dynamics simulations, could further improve functional discrimination and provide deeper insight into agonist- and antagonist-specific receptor interactions.

Data availability statement

Publicly available datasets were analyzed in this study. Ligand data were obtained from ChEMBL (https://www.ebi.ac.uk/chembl/), BindingDB (https://www.bindingdb.org/rwd/bind/index.jsp), and PubChem (https://pubchem.ncbi.nlm.nih.gov/).TRPM8 protein structures were obtained from the Protein Data Bank (PDB: https://www.rcsb.org/) with PDB IDs 7WRE and 9B6G.

Author contributions

NJ: Data curation, Visualization, Validation, Writing – original draft, Formal Analysis, Writing – review and editing, Investigation. PB: Writing – review and editing, Resources, Investigation, Funding acquisition, Writing – original draft, Formal Analysis, Supervision, Methodology, Conceptualization.

Funding

The author(s) declared that financial support was received for this work and/or its publication. We thankfully acknowledge the support of the Wolfson Foundation and the Royal Society for a

Royal Society Wolfson Fellowship (grant RSWF\R1\221005) and EPSRC (grant EP/X012026/1).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fddsv.2026.1770904/full#supplementary-material

References

Agarwal, R., and Smith, J. C. (2023). Speed vs accuracy: effect on ligand pose accuracy of varying box size and exhaustiveness in AutoDock vina. *Mol. Inf.* 42, e2200188. doi:10.1002/minf.202200188

Bianchini, G., Tomassetti, M., Lillini, S., Sirico, A., Bovolenta, S., Za, L., et al. (2021). Discovery of novel TRPM8 blockers suitable for the treatment of somatic and ocular painful conditions: a journey through p *K* a and LogD modulation. *J. Med. Chem.* 64, 16820–16837. doi:10.1021/acs.jmedchem.1c01647

Charifson, P. S., Corkery, J. J., Murcko, M. A., and Walters, W. P. (1999). Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 42, 5100–5109. doi:10.1021/jm990352k

Chang, M. W., Ayeni, C., Breuer, S., and Torbett, B. E. (2010). Virtual screening for HIV protease inhibitors: a comparison of autoDock 4 and vina. *PLoS One* 5, e11955. doi:10.1371/journal.pone.0011955

Guo, Q., Hernandez-Hernandez, S., and Ballester, P. J. (2024). "Scaffold splits overestimate virtual screening performance," 58–72. doi:10.1007/978-3-031-72359-9_5

Imrie, F., Bradley, A. R., and Deane, C. M. (2021). Generating property-matched decoy molecules using deep learning. *Bioinformatics* 37, 2134–2141. doi:10.1093/bioinformatics/btab080

Koes, D. R., Baumgartner, M. P., and Camacho, C. J. (2013). Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model* 53, 1893–1904. doi:10.1021/ci300604z

Liu, Y., Mikrani, R., He, Y., Faran Ashraf Baig, M. M., Abbas, M., Naveed, M., et al. (2020). TRPM8 channels: a review of distribution and clinical role. *Eur. J. Pharmacol.* 882, 173312. doi:10.1016/j.ejphar.2020.173312

Ochoa, S. V., Casas, Z., Albarracín, S. L., Sutachan, J. J., and Torres, Y. P. (2023). Therapeutic potential of TRPM8 channels in cancer treatment. *Front. Pharmacol.* 14. doi:10.3389/fphar.2023.1098448

Ramachandran, R., Hyun, E., Zhao, L., Lapointe, T. K., Chapman, K., Hirota, C. L., et al. (2013). TRPM8 activation attenuates inflammatory responses in mouse models of colitis. *Proc. Natl. Acad. Sci.* 110, 7476–7481. doi:10.1073/pnas.1217431110

Riniker, S., and Landrum, G. A. (2015). Better informed distance geometry: using what we know to improve conformation generation. *J. Chem. Inf. Model* 55, 2562–2574. doi:10.1021/acs.jcim.5b00654

Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., et al. (2014). rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* 10, e1003571. doi:10.1371/journal.pcbi.1003571

Scardino, V., Galarce, M. J., Mignone, M. E., and Cavasotto, C. N. (2025). Enhancing the reliability of integrated consensus strategies to boost docking-based screening campaigns using publicly available docking programs. *Mol. Inf.* 44, e2445. doi:10.1002/minf.2445

Shapovalov, M. V., and Dunbrack, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* 19, 844–858. doi:10.1016/j.str.2011.03.019

Tran-Nguyen, V.-K., Junaid, M., Simeon, S., and Ballester, P. J. (2023). A practical guide to machine-learning scoring for structure-based virtual screening. *Nat. Protoc.* 18, 3460–3511. doi:10.1038/s41596-023-00885-w

Trott, O., and Olson, A. J. (2010). AutoDock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334

Weyer, A. D., and Lehto, S. G. (2017). Development of TRPM8 antagonists to treat chronic pain and migraine. *Pharm. (Basel)* 10. doi:10.3390/ph10020037

Wu, Z., Peng, S., Huang, W., Zhang, Y., Liu, Y., Yu, X., et al. (2024). The role and function of TRPM8 in the digestive system. *Biomolecules* 14, 877. doi:10.3390/biom14070877

Xu, L., Han, Y., Chen, X., Aierken, A., Wen, H., Zheng, W., et al. (2020). Molecular mechanisms underlying menthol binding and activation of TRPM8 ion channel. *Nat. Commun.* 11. doi:10.1038/s41467-020-17582-x

Yin, Y., Zhang, F., Feng, S., Butay, K. J., Borgnia, M. J., Im, W., et al. (2022). Activation mechanism of the mouse cold-sensing TRPM8 channel by cooling agonist and PIP 2. *Science* 1979, 378. doi:10.1126/science.add1268

Zhao, C., Xie, Y., Xu, L., Ye, F., Xu, X., Yang, W., et al. (2022). Structures of a mammalian TRPM8 in closed state. *Nat. Commun.* 13, 3113. doi:10.1038/s41467-022-30919-y

Zhou, Y., Zhang, W., and Sun, F. (2025). Acoltremon: the first TRPM8 agonist approved for the treatment of dry eye disease. *Drug Discov. Ther.* 19, 01048–01211. doi:10.5582/ddt.2025.01048