Improving Reliability of the Virtual Screening Process for Human Dihydroorotate Dehydrogenase Enzyme by Combination of the Ensemble and Consensus Docking Approaches

Garri Chilingaryan (✉ g_chilingaryan@mb.sci.am)

Institute of Molecular Biology

Narek Abelyan

Russian-Armenian University

Arsen Sargsyan

Institute of Molecular Biology

Karen Nazaryan

Institute of Molecular Biology

Andre Serobian

Foundation for Armenian Science and Technology

Hovakim Zakaryan

Institute of Molecular Biology

Research Article

Keywords: virtual screening, dihydroorotate, CBP, enzyme

Posted Date: March 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-352349/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Improving reliability of the virtual screening process for human dihydroorotate dehydrogenase enzyme by combination of the ensemble and consensus docking approaches

Garri Chilingaryan^{1,*}, Narek Abelyan^{2,4}, Arsen Sargsyan¹, Karen Nazaryan¹, Andre Serobian⁴, and Hovakim Zakaryan^{1,3}

¹Institute of Molecular Biology of NAS RA, Yerevan, 0014, Armenia

²Russian-Armenian University, Yerevan, 0051, Armenia

³Denovo Sciences, Yerevan, 0033, Armenia

⁴Advanced Solutions Center, Foundation for Armenian Science and Technology, Yerevan, 0033, Armenia

*g_chilingaryan@mb.sci.am

ABSTRACT

The inconsistencies in the performance of the virtual screening (VS) process, depending on the used software and structural conformation of the protein, is a challenging issue in the drug design and discovery field. Varying performance, especially in terms of early recognition of the potential hit compounds, negatively affects the whole process and leads to unnecessary waste of the time and resources. Appropriate application of the ensemble docking and consensus-scoring approaches can significantly increase reliability of the VS results. Dihydroorotate dehydrogenase (DHODH) is a key enzyme in the pyrimidine biosynthesis pathway. It is considered as a valuable therapeutic target in cancer, autoimmune and viral diseases. Based on the conducted benchmark study and analysis of the effect of different combinations of the applied methods and approaches, here we suggested a structure-based virtual screening (SBVS) workflow that can be used to increase the reliability of VS.

Introduction

The hDHODH is a rate-limiting enzyme of the de novo pyrimidine synthesis pathway. The enzyme catalyzes the oxidation of dihydroorotate to orotate, which is essential for the production of uridine monophosphate^{1,2} (Fig. 1). Inhibition of hDHODH results in pyrimidine depletion, thereby starving the cell of the essential nucleotides required in the cell cycle. In humans, DHODH has been widely recognized as a promising target for the treatment of cancer and autoimmune diseases^{3,4}. Furthermore, inhibitors of hDHODH are of exclusive interest as host-targeting broad-spectrum antiviral agents against a wide variety of DNA and RNA viruses⁵⁻⁷. Given the medical significance, there are ongoing efforts towards the discovery and development of selective inhibitors for hDHODH.

Structure-based virtual screening is an established powerful computational approach that is extensively used in drug design and discovery due to its dramatic time-, cost- and labor-savings⁸. At the time of this publication, the Protein Data Bank⁹ (PDB) contained more than 60 crystal structures of hDHODH determined in the holo form, complexed with substrates, or inhibitors. The current structural data sufficiently allows researchers to conduct VS experiments in order to find new selective inhibitors for hDHODH¹⁰⁻¹². However, based on the binding pose prediction success rates in the recent benchmark study, hDHODH has been characterized as a “very hard” and challenging molecular target for docking¹³. Furthermore, a number of open-source academic and commercial molecular docking software packages show quite varied performances depending on protein family and also on crystal structure of the protein¹³⁻²⁰. In this context, the traditional approach in using a single software and structural conformation to screen a large database of molecules may lead to the questionable results.

Considering the therapeutic value of hDHODH, we have decided to develop a manual VS workflow, that integrates several known methods and approaches for the enhancement of VS efficiency, in particular ensemble and consensus docking strategies²¹⁻²⁹. To validate this newly developed approach, we conducted a benchmark study. While we developed and tested SBVS workflow for hDHODH, we also tested and discussed its general application and efficiency.

Figure 1. Schematic representation of de novo pyrimidine biosynthesis pathway and the 3D structure of hDHODH and the ubiquinone binding site.

Results and Discussion

Selection of the representative set of the hDHODH ubiquinone binding site structural conformations

Information on the interacting amino acid residues of the ubiquinone binding site of hDHODH (Fig. 2A) was gathered from the 38 selected crystal structures of hDHODH complexed with inhibitors (x-ray resolution < 2 Å). As a result of clusterization based on RMS deviation of the interacting amino acid residues of the ubiquinone binding site of hDHODH, four clusters were selected (cutoff 0.45Å) (Fig. 2B). For each cluster, a representative structure was selected (PDB IDs: 4IGH, 6GK0, 6J3C, 5ZF4).

Figure 2. Distribution of interacting amino acid residues of the ubiquinone binding site of hDHODH among 38 crystal structures in percentage (A) and their clusterization based on the RMSD (B).

Virtual screening

We performed virtual screening of the benchmarking library for four software against four selected representative crystal structures of the hDHODH ubiquinone binding site (16 runs, Table S2). Based on these separate 16 runs, the range of the values for AUC was from 0.62 to 0.84, for BEDROC 20 from 0.09 to 0.49 and for EFs (1 %) from 1.49 to 14.93, EF (3%) from 1 to 10.95, EF (5%) from 1.19 to 8.66, EF (10%) from 1.19 to 6.12 (Table S3). The obtained data confirmed that performance of the SBVS procedure highly depends on the selected molecular docking software and representative structure of the protein. Based on the average value among four used structural conformations of hDHODH, AutoDock Vina and ICM demonstrated close and significantly better results than the remaining two software, LeDock and rDock (Fig. 3).

Figure 3. Average performances of the AutoDock Vina, ICM LeDock and rDock among four structures. A) AUC, B) BEDROC 20, C) EF for 1, 3, 5 and 10 percent.

Taken together, obtained results showed that the performance of virtual screening may significantly vary depending on the structural conformation of hDHODH. In case of AutoDock Vina, ICM and LeDock screening against 4IGH and 6J3C crystal structures resulted in relatively better performance, compared to other structures, while for rDock - 5ZF4 was relatively better structure (Table S3).

Ensemble docking and consensus-scoring

Another important issue is multiple possible combinations of ensemble docking and consensus-scoring approaches and the difference in their effect on the improvement of the VS procedure. For the combination of ensemble docking and consensus-scoring approaches two scenarios were tested: 1) Docking scores were combined among four structures for separate software, then consensus-scoring was used among different software (Fig. 4) and conversely 2) A consensus-scoring approach is implemented among different software for separate structures, then normalized scores were combined among all structures (Fig. 5).

In the first case, a combination of ensemble docking approach based on the maximum docking scores (Fig. 4B) with the following consensus-scoring approach based on the average scores (Fig. 4D, “average”) demonstrated relatively better among all tested combinations of software. This combination showed the most stable positive effect among all tested different combinations of ensemble and consensus-scoring approaches (Fig. 4 and Fig.5).

In the second case, combinations based on the average score values among software (Fig. 5A) followed by combinations based on the maximum scores demonstrated better results than other combinations (Fig. 5C, “maximum”).

Among all possible combinations of software, the combination of the ICM and AutoDock Vina demonstrated the best results based on the all evaluated metrics (Fig. 4 and Fig. 5).

The effect of the different combinations of the ensemble and consensus-scoring approaches on the outcome of the virtual screening varied highly, depending on the way those approaches were combined, especially in the terms of enrichment factor. For the ICM and AutoDock Vina combinations, the range of the values of EF for 1, 3, 5 and 10 % are from 10.45 to 16.42, from 7.96 to 12.44, from 7.46 to 8.66 and from 5.22 to 6.42, respectively, while the differences of BEDROC 20 varies from

(A) Ensemble docking

AVERAGE	EF	1%	EF	3%	EF	5%	EF	10%	AUC	BEDROC	20
Vina	10.45	8.96	6.27	4.48	0.80	0.37					
ICM	8.96	7.96	7.46	5.22	0.79	0.41					
LeDock	2.99	1.99	2.39	2.84	0.76	0.18					
rDock	1.49	4.48	3.58	3.13	0.72	0.21					

(B) Ensemble docking

MAXIMUM	EF	1%	EF	3%	EF	5%	EF	10%	AUC	BEDROC	20
Vina	14.93	8.46	6.27	3.88	0.77	0.37					
ICM	11.94	9.95	8.96	6.12	0.84	0.48					
LeDock	4.48	3.98	3.58	3.28	0.77	0.24					
rDock	1.49	3.48	3.58	3.13	0.73	0.20					

(C) Consensus-scoring

AVERAGE	EF	1%	EF	3%	EF	5%	EF	10%	AUC	BEDROC	20
V - Vina	10.45	5.97	6.57	5.37	0.85	0.38					
I - ICM	5.97	4.98	5.37	4.48	0.83	0.32					
V + L + R	7.46	5.47	3.88	4.33	0.81	0.30					
V + I + R	11.94	8.96	7.46	5.67	0.85	0.44					
V + I + L	8.96	8.46	6.27	5.22	0.85	0.39					
L + R	2.98	3.48	3.28	3.13	0.77	0.21					
I + R	7.46	7.96	6.57	5.22	0.81	0.39					
I + L	7.46	4.98	5.37	4.48	0.83	0.32					
V + R	7.46	6.47	6.87	4.93	0.81	0.35					
V + L	7.46	5.47	5.37	4.33	0.80	0.31					
V + I	10.45	10.95	7.76	5.82	0.85	0.48					

MAXIMUM	EF	1%	EF	3%	EF	5%	EF	10%	AUC	BEDROC	20
V - Vina	4.48	5.47	5.07	5.22	0.85	0.33					
I - ICM	1.49	3.98	4.18	4.63	0.83	0.27					
V + L + R	4.48	4.98	4.48	4.18	0.79	0.29					
V + I + R	5.97	7.96	6.57	5.82	0.84	0.40					
V + I + L	5.97	5.97	5.07	4.63	0.85	0.34					
L + R	1.49	3.48	3.28	2.69	0.75	0.19					
I + R	1.49	5.47	6.27	5.22	0.80	0.35					
I + L	4.48	3.98	4.18	4.33	0.84	0.28					
V + R	5.97	6.97	7.46	4.78	0.79	0.35					
V + L	5.97	4.98	5.07	3.88	0.79	0.29					
V + I	11.94	7.96	7.76	6.42	0.85	0.45					

(D) Consensus-scoring

AVERAGE	EF	1%	EF	3%	EF	5%	EF	10%	AUC	BEDROC	20
V - Vina	13.43	9.95	8.36	5.52	0.85	0.47					
I - ICM	11.94	7.96	6.57	5.97	0.84	0.42					
V + L + R	11.94	7.46	5.37	4.18	0.80	0.35					
V + I + R	14.93	10.95	9.55	5.67	0.84	0.48					
V + I + L	13.43	10.95	8.06	5.52	0.85	0.48					
L + R	4.48	3.98	4.48	3.28	0.78	0.24					
I + R	4.48	7.96	8.36	5.97	0.83	0.43					
I + L	10.45	9.45	8.36	5.67	0.85	0.46					
V + R	7.46	6.47	5.67	4.18	0.79	0.34					
V + L	13.43	7.96	5.67	4.03	0.79	0.36					
V + I	14.93	12.44	8.66	5.52	0.85	0.50					

MAXIMUM	EF	1%	EF	3%	EF	5%	EF	10%	AUC	BEDROC	20
V - Vina	5.97	6.47	6.27	5.07	0.82	0.35					
I - ICM	2.99	5.97	5.37	5.37	0.83	0.33					
V + L + R	5.97	5.47	5.67	4.03	0.76	0.30					
V + I + R	4.48	6.97	7.46	5.52	0.82	0.40					
V + I + L	10.45	7.46	6.87	5.07	0.84	0.40					
L + R	2.99	3.48	3.58	3.13	0.76	0.22					
I + R	2.99	6.47	6.87	5.67	0.83	0.38					
I + L	8.96	6.47	6.87	5.52	0.85	0.39					
V + R	5.97	7.46	6.27	4.78	0.77	0.34					
V + L	8.96	6.47	5.67	3.88	0.77	0.32					
V + I	14.93	9.95	8.36	5.82	0.84	0.46					

Figure 4. AUC, BEDROC 20 and EFs of different combinations of consensus-scoring and ensemble approaches. Combination of docking scores among four structures for separate software, based on the average (A) and the maximum (B) docking score values. Consensus-scoring in case of different software combinations (C and D). Colors correspond to the heatmap of the EFs values among combinations.

<

Figure 5. AUC, BEDROC 20 and EFs of different combinations of consensus-scoring and ensemble approaches. Combination of docking scores among four software for separate structures, based on the average (A) and the maximum (B) docking score values. Ensemble docking among four structures in case of different software combinations (C and D). Colors correspond to the heatmap of the EFs values among combinations.

0.45 to 0.50 (Fig. 4 and Fig. 5, V+I). For the combination of all four software EFs values varied from 2.99 to

We tested two scenarios, using the selected structures, in combination and separately (Fig. 6).

Figure 6. Performance of the selected combinations of consensus-scoring and ensemble approaches for AutoDock Vina and ICM in case of using two best structures (6J3C and 4IGH) separately and in combination. A) AUC, B) BEDROC 20, C) EFs values.

As seen in figure 6, compared to the scenarios when all four structures were applied (Fig.4 and Fig. 5), elimination of undesirable structures has not significantly affected the performance and only in two cases values of EF (1 %) slightly increased: 1) when both 6J3C and 4IGH structures are used for both software and consensus-scoring approach based on the average score values followed by an ensemble approach where maximum score values are applied and 2) when only the 6J3C structure is considered for both software and consensus-scoring based on the average values are applied (Fig. 6). When only 6J3C structure is considered, values of EF (3 %) slightly increased for the first case and decreased for the second case (Fig. 6). Two points are needed to be noted here: 1) minor changes in the performance, after elimination of “undesirable” structures, are justified by the fact that previously applied ensemble and consensus-scoring approaches have already minimized the impact of those structures, and 2) these small differences, observed in cases of using different combinations of two best structures (6J3C and 4IGH), may be highly dependent on the set of active compounds of the dataset used and should not be taken into account for the actual VS procedure. This shows that elimination of “undesirable” structural conformations (Fig. 6) does not negatively affect the outcome of VS and linearly decreases computational time of VS procedure. It follows that, if it is possible to distinguish the reliability of the structures beforehand, it is reasonable to limit the selection of structural conformations to the most appropriate ones.

Consensus binding poses

As the ensemble approach (multiple structures) was implemented, it is reasonable to evaluate success rates of software to predict crystal conformations of all 67 active compounds using selected representative crystal structures of hDHODH (Table 1). Based on our results (Table 1), all software performed reasonably well, with LeDock being most consistent among 4 crystal structures. Therefore, in addition to the methods used, we decided to apply an approach described by Houston and Walkinshaw [25]. These authors recommended to use several docking software to identify consensus binding poses for compounds, which increases reliability of predicted poses and consequently improve hit rates.

We have tested the effect of the application of the consensus binding pose approach for the combination of all four software (Table 2) and combination of the AutoDock Vina and ICM (Table 3). In order to evaluate the effect of consensus binding

	Vina (%)	ICM (%)	rDock (%)	LeDock (%)
4IGH	82.08	85.07	79.1	82.08
5ZF4	76.12	82.08	79.1	82.08
6J3C	79.1	71.64	71.3	86.57
6GK0	71.64	79.1	77.61	77.61

Table 1. Success rates of binding pose predictions of crystal conformation of the 67 active compounds.

approach in case of using different combinations of structural conformations, same scores (scores of the combination of ensemble docking based on the maximum scores of the 6J3C and 4IGH structures for Vina and ICM, separately, and following consensus-scoring among that software based on the average score values) for compounds were applied.

Structures used for CBP	active/inactive	Percentage of active compounds in list (%)	Top 25	Top 50	Top100
Initial dataset (no CBP)	67/1933	3.35	14	24	31
4 structures	27/350	7.16	10	16	19
4IGH + 6J3C	30/457	6.16	12	19	21
6J3C	36/609	5.58	14	17	25
4IGH	40/758	5.01	14	19	26
5ZF4	37/745	4.73	14	17	25
6GK0	40/851	4.49	14	18	27

Table 2. Effect of the application of consensus binding pose filter in case of using combination of four used software (AutoDock Vina, ICM, LeDock, rDOCK).

Structures used for CBP	active/inactive	Percentage of active compounds in list (%)	Top 25	Top 50	Top100
Initial dataset (no CBP)	67/1933	3.35	14	24	31
4 structures	42/738	5.39	16	22	29
4IGH + 6J3C	49/989	4.72	17	24	33
6J3C	54/1236	4.19	13	22	28
4IGH	59/1413	4.01	15	24	30

Table 3. Effect of the application of consensus binding pose filter in case of using combination of the AutoDock Vina and ICM.

As seen in table 2 and table 3, inclusion of the additional software and structures affected the ratio of active to inactive compounds in a positive way for both cases. In case of using all available software for filtering with the CBP approach (Table 2), a number of active compounds in the top of the ranked list, along with inclusion of additional structures, noticeably decreases compared to the initial dataset. In contrast, by using the best combination of software the difference between using all four, two and one structure is less significant compared to the initial dataset and in some cases number of compounds actually increases (Table 3).

Therefore, if there is no pre-validated data on software and structures performance and one is not able to distinguish their selection to the most appropriate ones, then it is reasonable not to use multiple structures of protein for the filtering with the CBP approach. Despite the fact that use of multiple structures for the filtering with the CBP approach may increase the percentage of potential hit compounds, unregulated elimination of some active compounds, can negatively affect virtual screening procedure in terms of early recognition. However, when one is able to conduct a manual benchmark study for the investigated protein beforehand, or has pre-validated data on the performance and “suitability” of selected software and structural conformations of the protein, limitation of both software and structures to the most appropriate ones is advantageous. Alongside with the increase in the percentage of potential active compounds in the library, application of such a filter can positively affect VS procedure in terms of early recognition, by additional placement of the active compounds in the top of the ranked list.

The main effect of the application of the ensemble and consensus docking approaches is in the increase of the reliability of VS results. As seen in figure 7, even in the case when pre-validated data is not taken into account and all available software and structural conformations of protein are used, the outcome (number of active compounds in top of the list) is similar to that of the best “one software - one structure” performance. However, if pre-validated data is available application of those approaches can significantly improve the performance of VS procedure, especially in terms of early recognition of potential hit compounds

(Fig. 7).

Figure 7. Comparison of the number of active compounds in the top 25 (1.25%), 50 (2.5%) and 100 (5%) of the ranked lists after application of the best identified combinations of the ensemble and consensus approaches to the average number of compounds (based on the 16 separate runs) and best “one software - one structure” case (ICM and 4IGH).

Conclusion

Taking into the account inconsistencies in the performance of the VS process, depending on the software and structural conformations of protein, appropriate application of the ensemble and consensus docking approaches significantly increases reliability of the VS results. However, their inappropriate integration can lead to negative outcomes in terms of performance. Based on the conducted benchmark and analysis of the different combinations of the applied methods and approaches, here we suggested a SBVS workflow that can be used to increase the reliability of the VS procedure (Fig. 8). In cases, like ours, when one is able to use pre-validated data, and limit the selection of the software and structures to the most appropriate ones, use of such workflow can increase the performance of the VS procedure in terms of early recognition of the potential hit compounds, which is of paramount importance in real-world screening applications, where researchers test top-ranked molecules in biological assays due to their high costs. Therefore, pre-validation of docking software and protein structures, before the actual VS procedure, is highly recommended.

Methods

Structure selection and preparation

38 crystal structures of hDHODH in complex with inhibitors (resolution < 2 Å) have been selected from the Protein Data Bank (PDB) (Table S1). Ligands, water, metal ions and other co-crystallized molecules were removed, except co-factor of hDHODH - flavin mononucleotide (FMN). Then protein structures were superimposed using ICM-browser³⁰. For the selection of a representative set of structures, clusterization method based on root-mean-square deviation (RMSD) of the interacting amino acid residues (Cα atoms) of ubiquinone binding site of hDHODH was performed using bio3D R package³¹. Important amino acid residues of the ubiquinone binding site were selected based on the distribution of the interacting amino acids with the co-crystallized ligands among 38 used crystal structures (based on the annotations of crystal complexes) (Fig. 2A). Polar hydrogen atoms were added to the structures with the protonation state at pH = 7, using Open Babel 2.4.0 software³².

Benchmarking dataset and evaluation metrics

Directory of Useful Decoys: Enhanced (DUD-E)³³ dataset of decoys for the hDHODH was downloaded, filtered for duplicates and 1933 unique decoys were selected. 67 crystal structures of hDHODH in complexes with inhibitors, including 38 complexes used for the selection of target's representative structures, were downloaded from the PDB (Table S1)⁹. Co-crystallized ligands

Figure 8. Schematic representation of the developed SBVS workflow.

were extracted from corresponding complexes and used as active molecules for the benchmark study. Therefore, the final benchmarking library consisted of 67 active compounds and 1933 decoys (inactive compounds).

Several commonly used metrics, such as area under the ROC curve (AUC), Boltzmann-enhanced discrimination of ROC (BEDROC, with $\alpha = 20$), and enrichment factors (EF) for several percentage were used to evaluate and compare accuracy and performance of the conducted virtual screening procedures and calculated using rocker³⁴. Enrichment factor was calculated for 1, 3, 5 and 10. Since we applied a filter that changes the ratio of active to inactive compounds of the initial dataset and makes direct comparison of aforementioned metrics inaccurate, as an additional comparative metric we simply compared the numbers of active compounds found in the top 25, 50 and 100 of the ranked lists.

Docking software and parameters

Four molecular docking software packages were used, including three open-source/freely-distributed (AutoDock Vina³⁵, LeDock³⁶ and rDOCK³⁷) and one commercial (ICM³⁰) packages. Main differences in molecular docking software are search algorithm and scoring functions. LeDock uses evolutionary algorithm adopted in combination with simulated annealing search, which is used to generate the first generation of docking poses and physics and knowledge-based hybrid scoring function³⁶. RDock uses combination of stochastic and deterministic search techniques to generate low energy ligand poses and rDock master scoring function, which is a weighted sum of intermolecular, ligand intramolecular, site intramolecular, and external restraint terms³⁷. The ICM docking algorithm is a global energy optimization procedure based on Monte Carlo minimization³⁰. ICM has 2 scores: "ICM Score" and Potential of Mean Force (PMF) score. Only the ICM score, which is an empirical scoring function based on calculation of physiochemical properties of the receptor-ligand complex, was used in our study. AutoDock Vina uses "Iterated Local Search global optimizer" similar to that of ICM and Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method for the local optimization for generation of docking poses and hybrid scoring function (empirical + knowledge-based)³⁵.

The 67 crystal complexes of hDHODH with co-crystallized ligands were superimposed using icm-browser and docking box that covers all 67 ligands, based on their crystal structure conformations, and the amino acid residues of the ubiquinone binding site was used for AutoDock Vina and LeDock runs. In the case of the rDock, cavity mapping method based on the reference ligand was used. In the case of the ICM software, the docking grid was based on the interacting amino acid residues of the ubiquinone binding pocket of hDHODH (Fig. 2A). For all software packages, corresponding standard protocols were used with exception of rDOCK where the number of docking runs was set to 150.

Ensemble docking and consensus-scoring

In ensemble docking scores of each compound, obtained from docking runs using the same software but different structural conformers of a target protein, are combined using different methods²³.

In consensus-scoring scores of each compound, obtained from individual software for the same representative structure of protein, are combined using different methods²⁷. Since various docking software are used (i.e., different scoring functions), it is required to perform preceding score normalization. For the normalization of docking scores, from different software simple z-score function: $z = (x - \mu) / \sigma$, where x is a compound score, μ and σ - mean and standard deviation of scores of all compounds in the library, respectively, was used. For, ensemble docking and consensus-scoring approaches, docking scores were combined using two common data fusion methods: based on average and maximum (lowest binding energy) score values^{38,39}. We combined ensemble and consensus-scoring approaches by two alternative methods. 1) Docking scores of each compound obtained by one software for four structures were combined. Then, scores obtained by different software were

normalized and also combined. 2) Ddocking scores of compounds for each structural conformation of hDHODH obtained by different software were normalized and combined. Then, new normalized scores of each compound were combined among four structures.

Consensus binding pose approach (CBP)

An approach, based on consensus binding pose prediction and its use as a filter, described by Houston and Walkinshaw²⁶, was implemented. The difference in our case is use of multiple structural conformations of the target protein. After molecular docking runs, the top ranked docked conformation of each compound was selected. Those conformations were used for consensus binding poses evaluation. Two poses were regarded as similar (consensus) if the RMSD value was below 2 Å, which is a widely accepted value in docking predictions^{26,40,41}. All possible combinations of used docking programs were tested. Compounds, for which consensus binding poses for one structural conformation of protein were predicted by different software, passed into the new lists. Then those lists (one for each structure) were compared with each other, and if a compound was presented in all of the lists, then it was selected into the final filtered list.

References

1. Munier-Lehmann, H., Vidalain, P.-O., Tangy, F. & Janin, Y. L. On dihydroorotate dehydrogenases and their inhibitors and uses. *J Med Chem* **56**, 3148–3167, DOI: [10.1021/jm301848w](https://doi.org/10.1021/jm301848w) (2013).
2. Reis, R. A., Calil, F. A., Feliciano, P. R., Pinheiro, M. P. & Nonato, M. C. The dihydroorotate dehydrogenases: Past and present. *Arch. Biochem. Biophys.* **632**, 175–191, DOI: [10.1016/j.abb.2017.06.019](https://doi.org/10.1016/j.abb.2017.06.019) (2017).
3. Lolli, M. L. *et al.* Use of human Dihydroorotate Dehydrogenase (hDHODH) Inhibitors in Autoimmune Diseases and New Perspectives in Cancer Therapy. *Recent Pat Anticancer. Drug Discov* **13**, 86–105, DOI: [10.2174/1574892812666171108124218](https://doi.org/10.2174/1574892812666171108124218) (2018).
4. Madak, J. T., Bankhead, A., Cuthbertson, C. R., Showalter, H. D. & Neamati, N. Revisiting the role of dihydroorotate dehydrogenase as a therapeutic target for cancer. *Pharmacol. & Ther.* **195**, 111–131, DOI: [10.1016/j.pharmthera.2018.10.012](https://doi.org/10.1016/j.pharmthera.2018.10.012) (2019).
5. Hoffmann, H.-H., Kunz, A., Simon, V. A., Palese, P. & Shaw, M. L. Broad-spectrum antiviral that interferes with de novo pyrimidine biosynthesis. *Proc Natl Acad Sci U S A* **108**, 5777–5782, DOI: [10.1073/pnas.1101143108](https://doi.org/10.1073/pnas.1101143108) (2011).
6. Cheung, N. N. *et al.* Broad-spectrum inhibition of common respiratory RNA viruses by a pyrimidine synthesis inhibitor with involvement of the host antiviral response. *J Gen Virol* **98**, 946–954, DOI: [10.1099/jgv.0.000758](https://doi.org/10.1099/jgv.0.000758) (2017).
7. Hahn, F. *et al.* IMU-838, a Developmental DHODH Inhibitor in Phase II for Autoimmune Disease, Shows Anti-SARS-CoV-2 and Broad-Spectrum Antiviral Efficacy In Vitro. *Viruses* **12**, DOI: [10.3390/v12121394](https://doi.org/10.3390/v12121394) (2020).
8. Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M. & Taranto, A. G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front Chem* **8**, 343, DOI: [10.3389/fchem.2020.00343](https://doi.org/10.3389/fchem.2020.00343) (2020).
9. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242, DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235) (2000).
10. Fritzson, I. *et al.* Inhibition of human DHODH by 4-hydroxycoumarins, fenamic acids, and N-(alkylcarbonyl)anthranilic acids identified by structure-guided fragment selection. *ChemMedChem* **5**, 608–617, DOI: [10.1002/cmdc.200900454](https://doi.org/10.1002/cmdc.200900454) (2010).
11. Diao, Y. *et al.* Discovery of diverse human dihydroorotate dehydrogenase inhibitors as immunosuppressive agents by structure-based virtual screening. *J Med Chem* **55**, 8341–8349, DOI: [10.1021/jm300630p](https://doi.org/10.1021/jm300630p) (2012).
12. Xiong, R. *et al.* Novel and potent inhibitors targeting DHODH are broad-spectrum antivirals against RNA viruses including newly-emerged coronavirus SARS-CoV-2. *Protein Cell* **11**, 723–739, DOI: [10.1007/s13238-020-00768-w](https://doi.org/10.1007/s13238-020-00768-w) (2020).
13. Wierbowski, S. D., Wingert, B. M., Zheng, J. & Camacho, C. J. Cross-docking benchmark for automated pose and ranking prediction of ligand binding. *Protein Sci* **29**, 298–305, DOI: [10.1002/pro.3784](https://doi.org/10.1002/pro.3784) (2020).
14. Shamsara, J. Correlation between Virtual Screening Performance and Binding Site Descriptors of Protein Targets. *Int J Med Chem* **2018**, 3829307, DOI: [10.1155/2018/3829307](https://doi.org/10.1155/2018/3829307) (2018).
15. Chaput, L., Martinez-Sanz, J., Saettel, N. & Mouawad, L. Benchmark of four popular virtual screening programs: construction of the active/decoy dataset remains a major determinant of measured performance. *J Cheminform* **8**, 56, DOI: [10.1186/s13321-016-0167-x](https://doi.org/10.1186/s13321-016-0167-x) (2016).
16. Cross, J. B. *et al.* Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J Chem Inf Model.* **49**, 1455–1474, DOI: [10.1021/ci900056c](https://doi.org/10.1021/ci900056c) (2009).

17. Bolcato, G., Cuzzolin, A., Bissaro, M., Moro, S. & Sturlese, M. Can We Still Trust Docking Results? An Extension of the Applicability of DockBench on PDBbind Database. *Int J Mol Sci* **20**, DOI: [10.3390/ijms20143558](https://doi.org/10.3390/ijms20143558) (2019).
18. Li, X., Li, Y., Cheng, T., Liu, Z. & Wang, R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J. Comput. Chem.* **31**, 2109–2125, DOI: <https://doi.org/10.1002/jcc.21498> (2010).
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21498>.
19. Wang, Z. *et al.* Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **18**, 12964–12975, DOI: [10.1039/C6CP01555G](https://doi.org/10.1039/C6CP01555G) (2016). Publisher: The Royal Society of Chemistry.
20. Plewczynski, D., Łażniewski, M., Augustyniak, R. & Ginalska, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J Comput. Chem* **32**, 742–755, DOI: [10.1002/jcc.21643](https://doi.org/10.1002/jcc.21643) (2011).
21. Amaro, R. E. & Li, W. W. Emerging Methods for Ensemble-Based Virtual Screening. *Curr Top Med Chem* **10**, 3–13 (2010).
22. Fukunishi, Y. Structural ensemble in computational drug screening. *Expert. Opin Drug Metab Toxicol* **6**, 835–849, DOI: [10.1517/17425255.2010.486399](https://doi.org/10.1517/17425255.2010.486399) (2010).
23. Amaro, R. E. *et al.* Ensemble Docking in Drug Discovery. *Biophys J* **114**, 2271–2278, DOI: [10.1016/j.bpj.2018.02.038](https://doi.org/10.1016/j.bpj.2018.02.038) (2018).
24. Tian, S. *et al.* Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility. *J. Chem. Inf. Model.* **54**, 2664–2679, DOI: [10.1021/ci500414b](https://doi.org/10.1021/ci500414b) (2014). Publisher: American Chemical Society.
25. Li, Y. *et al.* Ensemble docking-based virtual screening toward identifying inhibitors against Wee1 kinase. *Futur. Medicinal Chem.* **11**, 1889–1906, DOI: [10.4155/fmc-2019-0022](https://doi.org/10.4155/fmc-2019-0022) (2019). Publisher: Future Science.
26. Houston, D. R. & Walkinshaw, M. D. Consensus docking: improving the reliability of docking in a virtual screening context. *J Chem Inf Model.* **53**, 384–390, DOI: [10.1021/ci300399w](https://doi.org/10.1021/ci300399w) (2013).
27. Palacio-Rodríguez, K., Lans, I., Cavasotto, C. N. & Cossio, P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci Rep* **9**, 5142, DOI: [10.1038/s41598-019-41594-3](https://doi.org/10.1038/s41598-019-41594-3) (2019).
28. Yang, J.-M., Chen, Y.-F., Shen, T.-W., Kristal, B. S. & Hsu, D. F. Consensus scoring criteria for improving enrichment in virtual screening. *J Chem Inf Model.* **45**, 1134–1146, DOI: [10.1021/ci050034w](https://doi.org/10.1021/ci050034w) (2005).
29. Craig, I. R., Essex, J. W. & Spiegel, K. Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *J Chem Inf Model.* **50**, 511–524, DOI: [10.1021/ci900407c](https://doi.org/10.1021/ci900407c) (2010).
30. Abagyan, R., Totrov, M. & Kuznetsov, D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. computational chemistry* **15**, 488–506 (1994). Publisher: Wiley Online Library.
31. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696, DOI: [10.1093/bioinformatics/btl461](https://doi.org/10.1093/bioinformatics/btl461) (2006).
32. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J Cheminform* **3**, 33, DOI: [10.1186/1758-2946-3-33](https://doi.org/10.1186/1758-2946-3-33) (2011).
33. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **55**, 6582–6594, DOI: [10.1021/jm300687e](https://doi.org/10.1021/jm300687e) (2012).
34. Lähti, S., Niinivehmas, S. & Pentikäinen, O. T. Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. *J Cheminform* **8**, 45, DOI: [10.1186/s13321-016-0158-y](https://doi.org/10.1186/s13321-016-0158-y) (2016).
35. Trott, O. & Olson, A. J. 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](https://doi.org/10.1002/jcc.21334) (2010).
36. Zhang, N. & Zhao, H. Enriching screening libraries with bioactive fragment space. *Bioorg Med Chem Lett* **26**, 3594–3597, DOI: [10.1016/j.bmcl.2016.06.013](https://doi.org/10.1016/j.bmcl.2016.06.013) (2016).
37. Ruiz-Carmona, S. *et al.* 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](https://doi.org/10.1371/journal.pcbi.1003571) (2014).
38. Willett, P. Combination of similarity rankings using data fusion. *J Chem Inf Model.* **53**, 1–10, DOI: [10.1021/ci300547g](https://doi.org/10.1021/ci300547g) (2013).
39. Bajusz, D., Rácz, A. & Héberger, K. Comparison of Data Fusion Methods as Consensus Scores for Ensemble Docking. *Molecules* **24**, DOI: [10.3390/molecules24152690](https://doi.org/10.3390/molecules24152690) (2019).

40. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **295**, 337–356, DOI: [10.1006/jmbi.1999.3371](https://doi.org/10.1006/jmbi.1999.3371) (2000).
41. Mena-Ulecia, K., Tiznado, W. & Caballero, J. Study of the Differential Activity of Thrombin Inhibitors Using Docking, QSAR, Molecular Dynamics, and MM-GBSA. *PLoS One* **10**, e0142774, DOI: [10.1371/journal.pone.0142774](https://doi.org/10.1371/journal.pone.0142774) (2015).

Acknowledgements

This work was supported by Faculty Research Funding Program (FRFP 2020) implemented by Enterprise Incubator Foundation (EIF) with support of Philip Morris International (PMI) Science. We thank Prof. Ruben Abagyan from the University of California, San Diego (USA) for the opportunity to use ICM-PRO.

Author contributions statement

G.C. and H.Z. designed and directed the study. G.C., N.A. and Ar.S. performed the computations and analysed the results. Both K.N. and An.S. contributed to the discussion of the results. All authors revised and approved the final manuscript.

Additional information

The authors declare no competing interests.

Figures

Figure 1

Schematic representation of de novo pyrimidine biosynthesis pathway and the 3D structure of hDHODH and the ubiquinone binding site.

Figure 2

Distribution of interacting amino acid residues of the ubiquinone binding site of hDHODH among 38 crystal structures in percentage (A) and their clusterization based on the RMSD (B).

Figure 3

Average performances of the AutoDock Vina, ICM LeDock and rDock among four structures. A) AUC, B) BEDROC 20, C) EF for 1, 3, 5 and 10 percent.

A) Ensemble docking									
AVERAGE	EF 1%	EF 3%	EF 5%	EF 10%	AUC	BEDROC 20			
Vina	10.45	8.96	6.27	4.48	0.80	0.37			
ICM	8.96	7.96	7.46	5.22	0.79	0.41			
LeDock	2.99	1.99	2.39	2.84	0.76	0.18			
rDock	1.49	4.48	3.58	3.13	0.72	0.21			

B) Ensemble docking									
MAXIMUM	EF 1%	EF 3%	EF 5%	EF 10%	AUC	BEDROC 20			
Vina	14.93	8.46	6.27	3.88	0.77	0.37			
ICM	11.94	9.95	8.96	6.12	0.84	0.48			
LeDock	4.48	3.98	3.58	3.28	0.77	0.24			
rDock	1.49	3.48	3.58	3.13	0.73	0.20			

C) Consensus-scoring													
V - Vina I - ICM L - LeDock R - rDock	AVERAGE						MAXIMUM						
	EF 1%	EF 3%	EF 5%	EF 10%	AUC	BEDROC 20	EF 1%	EF 3%	EF 5%	EF 10%	AUC	BEDROC 20	
V+I+L+R	10.45	5.97	6.57	5.37	0.85	0.38	4.48	5.47	5.07	5.22	0.85	0.33	
I+L+R	5.97	4.98	5.37	4.48	0.83	0.32	1.49	3.98	4.18	4.63	0.83	0.27	
V+L+R	7.46	5.47	3.88	4.33	0.81	0.30	4.48	4.98	4.48	4.18	0.79	0.29	
V+I+R	11.94	8.96	7.46	5.67	0.85	0.44	5.97	7.96	6.57	5.82	0.84	0.40	
V+I+L	8.96	8.46	6.27	5.22	0.85	0.39	5.97	5.97	5.07	4.63	0.85	0.34	
L+R	2.98	3.48	3.28	3.13	0.77	0.21	1.49	3.48	3.28	2.69	0.75	0.19	
I+R	7.46	7.96	6.57	5.22	0.81	0.39	1.49	5.47	6.27	5.22	0.80	0.35	
I+L	7.46	4.98	5.37	4.48	0.83	0.32	4.48	3.98	4.18	4.33	0.84	0.28	
V+R	7.46	6.47	6.87	4.93	0.81	0.35	5.97	6.97	7.46	4.78	0.79	0.35	
V+L	7.46	5.47	5.37	4.33	0.80	0.31	5.97	4.98	5.07	3.88	0.79	0.29	
V+I	10.45	10.95	7.76	5.82	0.85	0.48	11.94	7.96	7.76	6.42	0.85	0.45	

D) Consensus-scoring													
V - Vina I - ICM L - LeDock R - rDock	AVERAGE						MAXIMUM						
	EF 1%	EF 3%	EF 5%	EF 10%	AUC	BEDROC 20	EF 1%	EF 3%	EF 5%	EF 10%	AUC	BEDROC 20	
V+I+L+R	13.43	9.95	8.36	5.52	0.85	0.47	5.97	6.47	6.27	5.07	0.82	0.35	
I+L+R	11.94	7.96	6.57	5.97	0.84	0.42	2.99	5.97	5.37	5.37	0.83	0.33	
V+L+R	11.94	7.46	5.37	4.18	0.80	0.35	5.97	5.47	5.67	4.03	0.76	0.30	
V+I+R	14.93	10.95	9.55	5.67	0.84	0.48	4.48	6.97	7.46	5.52	0.82	0.40	
V+I+L	13.43	10.95	8.06	5.52	0.85	0.48	10.45	7.46	6.87	5.07	0.84	0.40	
L+R	4.48	3.98	4.48	3.28	0.78	0.24	2.99	3.48	3.58	3.13	0.76	0.22	
I+R	4.48	7.96	8.36	5.97	0.83	0.43	2.99	6.47	6.87	5.67	0.83	0.38	
I+L	10.45	9.45	8.36	5.67	0.85	0.46	8.96	6.47	6.87	5.52	0.85	0.39	
V+R	7.46	6.47	5.67	4.18	0.79	0.34	5.97	7.46	6.27	4.78	0.77	0.34	
V+L	13.43	7.96	5.67	4.03	0.79	0.36	8.96	6.47	5.67	3.88	0.77	0.32	
V+I	14.93	12.44	8.66	5.52	0.85	0.50	14.93	9.95	8.36	5.82	0.84	0.46	

Figure 4

AUC, BEDROC 20 and EFs of different combinations of consensus-scoring and ensemble approaches. Combination of docking scores among four structures for separate software, based on the average (A) and the maximum (B) docking score values. Consensus-scoring in case of different software combinations (C and D). Colors correspond to the heatmap of the EFs values among combinations.

Figure 5

AUC, BEDROC 20 and EFs of different combinations of consensus-scoring and ensemble approaches. Combination of docking scores among four software for separate structures, based on the average (A) and the maximum (B) docking score values. Ensemble docking among four structures in case of different software combinations (C and D). Colors correspond to the heatmap of the EFs values among combinations.

Figure 6

Performance of the selected combinations of consensus-scoring and ensemble approaches for AutoDock Vina and ICM in case of using two best structures (6J3C and 4IGH) separately and in combination. A) AUC, B) BEDROC 20, C) EFs values.

Figure 7

Comparison of the number of active compounds in the top 25 (1.25%), 50 (2.5%) and 100 (5%) of the ranked lists after application of the best identified combinations of the ensemble and consensus approaches to the average number of compounds (based on the 16 separate runs) and best “one software - one structure” case (ICM and 4IGH).

Figure 8

Schematic representation of the developed SBVS workflow.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xls](#)
- [TableS2.xls](#)
- [TableS3.xls](#)