In silico screening for identifying compounds specifically bound to either NAD- or NADH-bound structures of redox-sensing transcriptional factor, Rex1

Chaeyoung Lim[1], Gyeongeun Lee[2] and Minho Lee[1,2,*]

[1]Department of Life Science; [2]Department of Industrial Pharmacy, Dongguk University-Seoul, Korea

Abstract

Recently, one of a Redox-sensing transcriptional factor, Rex1's three-dimensional structures were identified. Not only apo form of the structure but NAD+ and NADH-bound forms (PDB ID 5zz6 and 5zz7) were also deposited in the PDB. Although NAD+ and NADH share similar structures, Rex1 structures they bound were somewhat different. The purpose of this work is to identify compounds which can be bound to Rex1 NAD+-like or NADH-like. We applied two different strategies: one was based on molecular docking; the other was based on similarity searching.
In the first approach, we screened more than 100k compounds in ZINC database by docking those molecules into three binding sites in 5zz6 (NAD+ bound form) and two binding sites in 5zz7 (NADH bound form). Then we compared the predicted binding affinities to identify compounds, which is likely to specifically bind to either one structure. In the second approach, 31,484 compounds similar to either NAD or NADH were obtained based on PubChem fingerprint. Then, compounds similar to NAD but not similar to NADH (total 141) and those not similar to NAD but similar to NADH (total 106) were found, and Tanimoto coefficients with NAD and NADH were calculated. As a result, we identified five compounds which is likely be selectively bound to NADH-bound form of Rex1.

Datasets

PDB data
→ 5zz6, 5zz7

ZINC data
→ Provide commercially available compounds (in a ready dockable 3D format) for virtual screening.
The downloaded ZINC file must eventually be converted to ".pdbqt" format.

Method1 In this study, docking was performed using about 100,000 ZINC compounds.

Method2

PubChem
→ To find similar substances of NAD and NADH

SDF files
→ Obtain SDF files of NAD and NADH-like substances containing chemical information such as molecular formula, SMILES, 2D or 3D coordinates.

Tanimoto Coefficient
→ As one of the methods for measuring drug similarity, it is calculated using fingerprint, information that can identify molecular properties or compound similarity.

Affinity
→ Calculate the affinity between the target substance and NAD or NADH.

Figure1. To find similar substances in NAD and NADH, we obtained SDF files by PubChem Search with setting the tanimoto similarity of NAD and NADH. The SDF files contains 2D or 3D coordinates and chemical information such as molecular formula and SMILES. The Tanimoto coefficient used in this study is one of the methods to measure drug similarity.

Methods

Input data - ZINC.pdbqt → Docking → Primary data - 5zz6, 5zz7, rmHETATM.pdbqt, config files

↓

Check some samples using PyMOL

↓

Obtain the best model's affinity values by each ligand

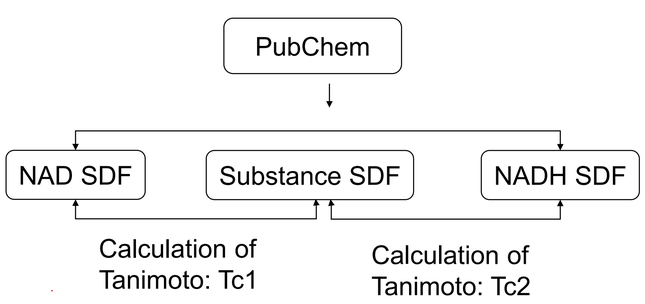
↓

Analysis

Figure2. Process
(1) Create rmHETATM -> Convert to pdbqt format by Autodock tool
(Deletion of hetero atom, Check that the hetero atom is removed with PyMOL)
(2) Create config files (Files contain calculated center coordinates and size of the specific ligand)
(3) Prepare the ligand.pdbqt (Use ZINC files instead of NAD and NADH)
*Format : .sdf->.mol2->.pdbqt
(4) Dock into Protein using Vina
Insert rmHETATM.pdbqt, config files, ligand.pdbqt into Vina program.
(5) Check some samples using PyMOL
(6) Using Vina result, obtain the best model's affinity values by ligand.
(7) Analysis
*rmHETATM: Models with removed ligands

Method 1

Method 2



Selection criteria

- Tc1 > 0.85 or Tc2 > 0.85

Figure3. Graphical representation describing two methods we used.

Results 1

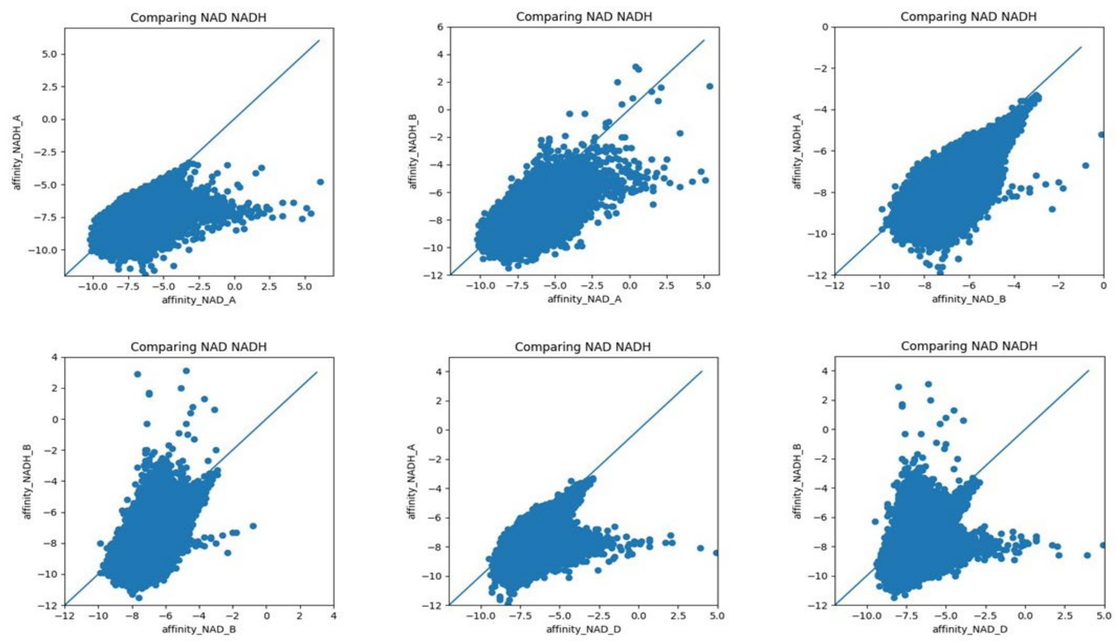


Figure4. NAD is at three positions in the reference protein model and is named as A, B and D, respectively. NADH is located at two positions in the reference protein model, and was arbitrarily set to A and B, respectively. The docking results of ZINC data were analyzed with the purpose of filtering out substances with relatively high affinity for only one of NAD and NADH. Referring to the scatter plot, a molecule with a high affinity value from only one side is a molecule to be filtered out. Additionally, there is a tendency according to the model docked at specific positions of NAD and NADH. Finally we identified five compounds which is likely be selectively bound to NADH-bound form of Rex1

Results 2

Based on the method2, we selected either NAD-like or NADH-like compounds. The molecules were then docked into the binding sites of structures of NAD-bound and NADH-bound sites. We finally selected two compounds (PubChem CID: 101043025 and 196369), which were predicted as more likely to bind to NADH-binding sites. Similarity values of the compounds to NAD and NADH are shown in Table 1 and predicted affinities are shown in Table 2.

Table 1. Tanimoto Coefficients with NADH and NAD for each substance

Substance(CID)	NADH tanimoto coefficient	NAD tanimoto coefficient
101043025	0.962352941	0.68487395
196369	0.895089286	0.695020747

Table 2. Predicted binding affinities at each site of NADH and NAD

Substance(CID)	Affinity NADH_A	Affinity NADH_B	Affinity NAD_A	Affinity NAD_B	Affinity NAD_D
101043025	-5.3	-5.5	-5.1	-4.4	-4.2
196369	-10.6	-11	-10.5	-9.4	-8.9

To validate the two compounds, we repeated docking calculation by increasing exhaustiveness option (reflecting docking accuracy) in Vina from default value (8) to 40 (Table 3) and 100 (Table 4). Unfortunately, all predicted binding affinities to NADH-binding sites were not stronger than those to NAD-binding sites.

Table 3. Predicted binding affinities to NADH and NAD binding sites for exhaustiveness 40

Substance(CID)	Affinity NADH_A	Affinity NADH_B	Affinity NAD_A	Affinity NAD_B	Affinity NAD_D
101043025	-5.1	-4.2	-4.2	-5.3	-5.5
196369	-10.9	-9.4	-9	-10.7	-11

Table 4. Predicted binding affinities to NADH and NAD binding sites for exhaustiveness 100

Substance(CID)	Affinity NADH_A	Affinity NADH_B	Affinity NAD_A	Affinity NAD_B	Affinity NAD_D
101043025	-5.1	-4.4	-4.2	-5.3	-5.5
196369	-10.9	-9.5	-9	-10.7	-11

Preliminary conclusion

Through this experiment, we could find five compounds suitable for the purpose. By contrast, based on the second method using similarity searching, we failed to identify appropriate compounds. This may due to large difference in the number of molecules screened. The results have limitation in that they are based on computational simulation, so additional experimental validation should be needed in subsequent researches.