

# **Protocol for fast screening of multi-target drug candidates: Application to Alzheimer's disease**

Nguyen Quoc Thai<sup>1,2,3</sup>, Nguyen Hoang Linh<sup>1</sup>, Huynh Quang Linh<sup>3</sup>, and Mai Suan Li<sup>1,4</sup>

<sup>1</sup>*Institute for Computational Sciences and Technology, SBI building, Quang Trung Software  
City, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam*

<sup>2</sup>*Dong Thap University, 783 Pham Huu Lau Street, Ward 6, Cao Lanh City, Dong Thap, Vietnam*

<sup>3</sup>*Biomedical Engineering Department, University of Technology -VNU HCM, 268 Ly Thuong  
Kiet Str., Distr. 10, Ho Chi Minh City, Vietnam*

<sup>4</sup>*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*

**Corresponding Author** Mai Suan Li, [masli@ifpan.edu.pl](mailto:masli@ifpan.edu.pl)

## \Abstract

The treatment of many diseases may require drugs that are capable to attack multiple targets simultaneously. Obviously, the virtual screening of multi-target drug candidates is much more time consuming compared to the single-target case. This, in particular, concerns the last step of virtual screening where the binding free energy is computed by conventional molecular dynamics simulation. To overcome this difficulty we propose a simple protocol which is relied on the fast steered molecular dynamics simulation and on available experimental data on binding affinity of reference ligand to a given target. Namely, first we compute non-equilibrium works generated during pulling ligands from the binding site using the steered molecular dynamics method. Then as top leads we choose only those compounds that have the non-equilibrium work larger than that of a reference compound for which the binding free energy has been already known from experiment.

Despite many efforts no cures for AD (Alzheimer's disease) have been found. One of possible reasons for this failure is that drug candidates were developed for a single target, while there are exist many possible pathways to AD. Applying our new protocol to five targets including amyloid beta fibril, peroxisome proliferator-activated receptor  $\gamma$ , retinoic X receptor  $\alpha$ ,  $\beta$ - and  $\gamma$ -secretases, we have found two potential drugs (CID 16040294 and CID 9998128) for AD from the large PubChem database. We have also shown that these two ligands can interfere with the activity of popular Acetylcholinesterase target through strong binding towards it.

**Key words** Multi-target drug design, Steered molecular dynamics, Alzheimer's disease, Amyloid beta peptide,  $\beta$ -secretase, PPAR $\gamma$ , LXR $\alpha$

## Introduction

Nowadays, the computer is often used for identifying drug candidates. The computer-aided drug design involves multiple steps such as the application of Lipinski's rule [1], molecular docking and molecular dynamics simulation to estimate the receptor-ligand binding free energy ( $\Delta G_{\text{bind}}$ ). The last step is the most time consuming. For estimation of  $\Delta G_{\text{bind}}$  one can use different methods [2-5] including the MM/PBSA (molecular mechanics Poisson-Boltzmann surface area) method [6] which are more reliable than the existing docking methods but, in return, are very time consuming. Recently it has been recognized that the steered molecular dynamics (SMD) method [7-9] is as accurate as the MM/PBSA method in estimating relative  $\Delta G_{\text{bind}}$  but CPU much less demanding [10-12]. In SMD the binding affinity is characterized either by rupture force  $F_{\text{max}}$  occurring in the force-time/extension curve or by non-equilibrium work  $W_{\text{pull}}$ , generated during ligand escape from the binding pocket. Note that, compared to  $F_{\text{max}}$ ,  $W_{\text{pull}}$  was shown to have a better correlation with experimental inhibition constant ( $K_i$ ) [13]. Therefore it will be used throughout this paper.

Because the multi-target drug design requests more computation time than the single-target case we propose to apply the SMD instead of MD simulation in the last step of virtual screening. However, the SMD is able to predict relative but not absolute binding affinities making the choice of top compounds uncertain. To handle with this difficulty we offer a simple trick which combines SMD results with available experimental data on the inhibition constant of a reference compound which is a good binder to a given target. Namely we choose only those compounds that have  $F_{\text{max}}$  or  $W_{\text{pull}}$  larger than that of the reference compound. For illustration new approach will be applied to seek potential drugs that can attack several targets of the Alzheimer's disease (AD).

AD was first identified as a main culprit causing death of people with mental illness [14]. AD patient suffers from the cognitive decline associated with the neurons loss [15]. The etiology of this disease remains obscure and there are dozens hypotheses about its causes that are roughly grouped into three major categories [16]: cellular, genetic, and molecular imbalances. Consequently, various potential drugs have been developed based on distinct hypotheses. According to the oldest cholinergic hypothesis [17] AD occurs due to a reduced level of neurotransmitter acetylcholine. Currently available on market drugs are AChE (acetylcholinesterase) inhibitors, which have been developed basing on this hypothesis.

However, recent experimental evidences strongly support the so called amyloid cascade hypothesis [15] (molecular imbalance) positing that the AD etiology is associated with self-assembly of amyloid beta ( $A\beta$ ) peptides inside the brain. A lot of works have been done on designing inhibitors for either  $A\beta$  aggregation or clearance of  $A\beta$  fibrils/oligomers as potential drugs [18-20]. Note that transient oligomers are probably more toxic than mature fibrils [21, 22], although the relationship between toxic and other  $A\beta$  forms remains obscure [23]. Another

possible A $\beta$  therapy is based on preventing A $\beta$  production. Because A $\beta$  peptides are products of proteolytic cleavage of APP (amyloid precursor protein) by  $\beta$ -secretases (BACE-1) and  $\gamma$ -secretases, different compounds inhibiting BACE-1 [24] and  $\gamma$ -secretase [25] activity have been developed .

In terms of genetic imbalance, many genes were shown to level up AD risk, but the gene of Apolipoprotein E (APOE) [26, 27], which was expressed by nuclear receptor proteins [28], is one of the most important. In humans APOE has 3 allelic variants APOE2, APOE3, and APOE4, but only APOE4 is considerably overexpressed in AD patients [29]. Peroxisome proliferator activated receptors have three types [28] (alpha, beta and delta) are denoted as PPAR $\alpha$ , PPAR $\beta$  and PPAR $\gamma$  . Among them PPAR $\gamma$  is overexpressed in the AD patient brain [30] and is appreciated to have ability to modulate inflammation response [31] in animal model of AD. PPAR $\gamma$  agonist reduced not only A $\beta$  plaque burden, but also inflammation [32, 33]. Thus PPAR $\gamma$  has emerged as therapeutic targets for AD [34].

LXR (the liver X receptor), which has two isoforms LXR $\alpha$  and LXR $\beta$ , is in close relation with PPARs. Treatment with LXR $\alpha$  agonist T090131, for instance, reduced A $\beta$  production in an AD mouse model [35]. However, T090131 was found to enhance plasma as well as liver triglycerides, suggesting that this compound is not a good drug candidate. From this perspective the development of prominent LXR agonists with no side effects is of clinical importance. Thus we will consider LXR $\alpha$  as a target for AD therapy [34].

Because of existing various mechanisms and multifactorial progression of AD, multi-target drug design has occurred as an innovative therapeutic approach for AD [36, 37]. This novel strategy has been recently focused on bi-target cases [38-41]. A number of carbazole compounds, which bind to AChE and exhibit modest A $\beta$  fibril formation but display low affinities for BACE-1, were developed [42].

Recently, Prati *et al* [43] have provided a summary on development of AD multi-target drug candidates using framework combination [44] and fragment-based drug discovery (FBDD) strategies for novel MTDL candidates against AD. Memantine-Galantamine hybrid, for example, can simultaneously bind to AChE (IC<sub>50</sub> = 1.16 nM) and N-methyl-D-aspartate receptor (IC<sub>50</sub> = 4.6  $\mu$ M). The FBDD strategy become popular among researchers as an optimal starting point for multi-target drug design [45, 46]. This strategy yielded 6-amino-4-phenyl-3,4-dihydro-1,3,5-triazin-2(1H) compound , which inhibited BACE-1 and glycogen synthase kinase-3 (GSK-3 $\beta$ ) enzymes [47].

No attempt, however, has been made to design compounds that can hit A $\beta$  cascade hypothesis targets as well as targets related to genetic imbalance. Here we try to solve the multi-target problem by seeking AD potential drugs, which would have good binding affinity to 5 targets including amyloid beta fibril,  $\beta$ -secretase,  $\gamma$ -secretase, PPAR $\gamma$  and LXR $\alpha$ . Contrary to prior works, where novel candidates were designed by combining structures of relevant single-target

leads, we searched drug candidates from PubChem database (see <https://pubchem.ncbi.nlm.nih.gov/>) which contains more than 1.4 million compounds. In the multi-step virtual screening we have employed the Lipinski's rule, docking simulations and PreADMET to obtain top compounds that can easily cross the BBB (blood brain barrier). Finally our novel approach, which combines SMD with experimental data on binding affinity of reference compounds, was used to reveal multi-target leads. We predict that two compounds CID 16040294 and CID 9998128 are the best candidates for five mentioned above targets of AD. In addition, using the MMPB-SA method we have shown that these compounds also have good binding affinity towards AChE.

## Material and methods

### *Ligands and Lipinski's rules*

Around 1.4 million compounds were taken from the PubChem [47]. We used Lipinski's rule [48] to filter drug-like compounds. From the whole set we obtained only 5372 ligands (Fig. 1) for further study.

### *Multiple targets (receptors)*

There are several experimental structures of A $\beta$  protofibrils that can be used for AD drug design [49], but we have chosen the crystal structure, obtained by solid state NMR for fragment A $\beta_{11-42}$  with PDB code 2MXU [50]. The rationale for our choice is that this structure is likely reliable containing two turns in each chain [51]. The experimentally resolved structure of PPAR $\gamma$ , which is available in PDB under code 4EMA [52], has the ligand binding domain at residues 235-505. The PDB structure 4K6I [53] of retinoic X receptor alpha (RXR $\alpha$ ) was used for simulation.

The structure of the intramembrane protease  $\gamma$ -secretase was resolved experimentally [54], but its binding site has not been determined yet. Therefore, for studying ligand binding to  $\gamma$ -secretase we employed the structure of complex structure of presenilin homologue (PSH) which has the PDB ID 4Y6K [55]. We have made this choice because (i) PSH share high structural similarity with component presenilin PS1 of  $\gamma$ -secretase [55, 56]; (ii) PSH and human  $\gamma$ -secretase exhibit similar cleavage propensities towards APP C99 [54] and (iii) the A $\beta_{42}$ :A $\beta_{40}$  ratio of cleavage by PSH is almost the same as that by  $\gamma$ -secretase [55].

For  $\beta$ -secretase, the structure with PDB ID 1M4H [57] was used for simulation. To test the binding affinity of predicted top leads to AChE the PDB structure 1H22 was taken. The structures of six receptors are shown in SI (Supporting Information) (Figure S1).

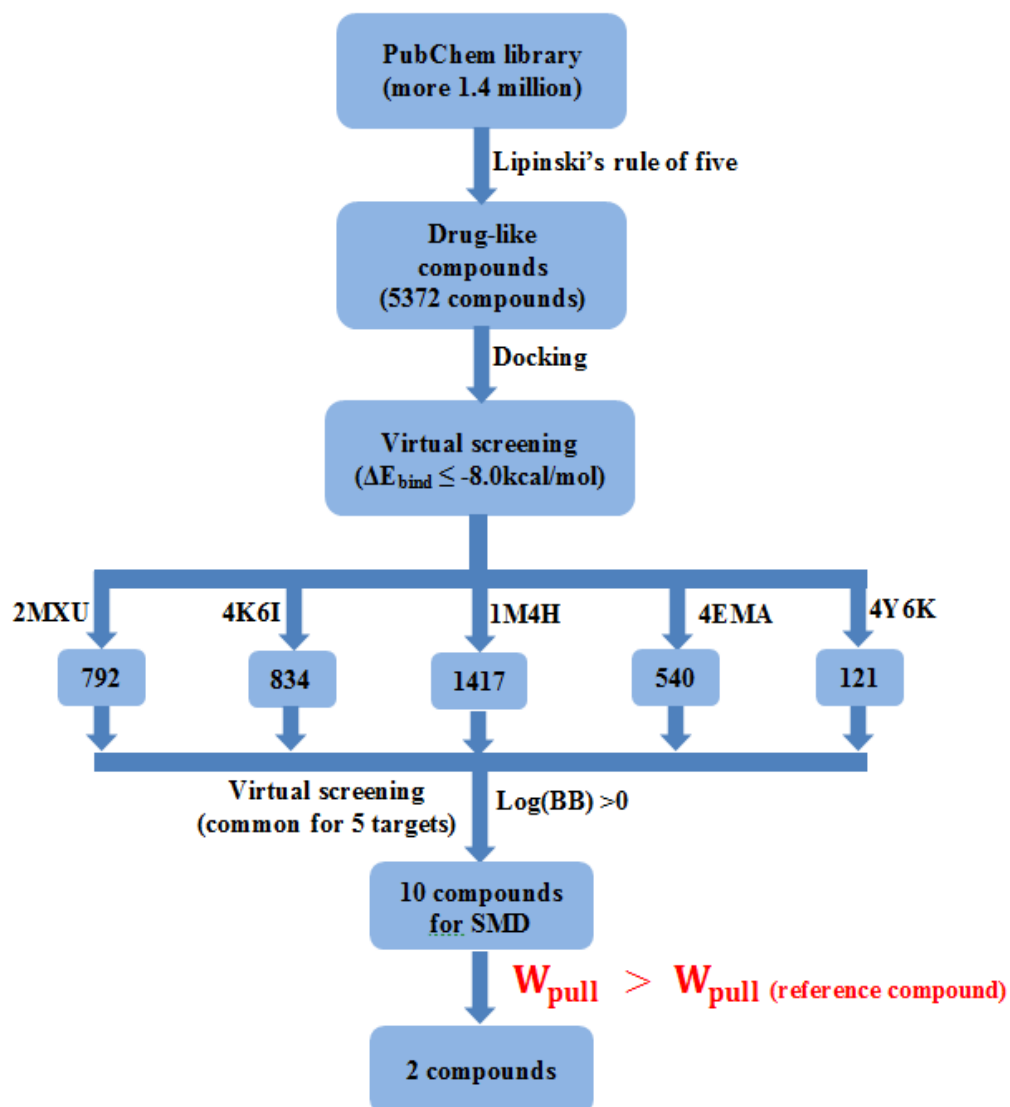


Figure 1. Multi-step screening procedure. Our new step involves the application of SMD to seek for compounds that have the non-equilibrium work larger than that of the reference compound for a given target.

#### *Reference compounds*

In our new approach top hits are selected by comparing the non-equilibrium works with that of a reference compound for which the experimental data on inhibition constant are available. Table 1 shows the list of reference compounds for six targets with their binding free energy measured experimentally and  $\Delta E_{\text{bind}}$  calculated by the docking method. Their 2D chemical structures of reference compound are shown in Figure S2 in SI.

Table 1. Experimental inhibition constants of reference compounds for six targets. The binding free energy was estimated using the formula  $\Delta G_{\text{exp}} = RT\ln(K_i)$ , where  $R=1.987\times 10^{-3}$  kcal/mol,  $T=300$  K, and  $K_i$  is measured in M. The binding energy obtained by the docking simulation is also displayed for comparison.

Target	Ligand	K <sub>i</sub> (nM)	$\Delta G_{\text{exp}}$ (kcal/mol)	docking $\Delta E_{\text{bind}}$ (kcal/mol)	Ref.
A $\beta$ protofibril (2MXU)	Curcumin	0.2	-13.3	-7.2	F.S. Yang et al [58]
$\beta$ -secretase (1M4H)	71548079	0.03	-14.4	-8.9	A.F. Abdel-Magid [59]
$\gamma$ -secretase (4Y6K)	44354431	0.49	-12.7	-8.5	A.P. Owens [60]
PPAR $\gamma$ (4EMA)	18944089	0.4	-12.9	-8.3	D.R.Buckle [61]
RXR $\alpha$ (4K6I)	Bexarotene	14	-10.8	-12.6	L.J. Farmer et al [62]
AChE (1H22)	10107976	0.03	-14.4	-9.5	P. Camps et al [63]

### *Docking method*

As in the single target case, we have prepared PDBQT file using Autodock Tool 1.5.4 [64]. The docking simulation has been carried out by Autodock Vina version 1.1 [65]. To achieve reliable results in global search we set exhaustiveness parameter equal 600. Because the binding pocket of A $\beta_{42}$  fibril is not known to cover the whole fibril the 7.4x4.6x5.6 nm box was chosen to cover the entire protofibril. For other targets whose binding sites were known from experiment, the boxes were chosen to cover just the binding sites with grid dimensions 4.0x4.0x4.0 nm (1M4H), 1.6x1.6x1.4 nm (4Y6K), 2.0x1.6x1.6 nm (4K6I), and 1.6x1.8x2.0 nm (4EMA). In docking

simulation the receptor dynamics was omitted. The scoring function is the binding (lowest) energy  $\Delta E_{\text{bind}}$  obtained in the best docking mode.

### *Molecular Dynamic (MD) Simulation*

The MD simulation was performed using the force field AMBER-f99SB-ILDN [66] combined with TIP3P [67]. Force field parameters for small compounds were calculated using Antechamber [68] and Acpype [69] basing on the General Amber Force Field (GAFF) [70]. The AM1-BCC [71] charge model was used to calculate the atomic point charges. Parameters used for simulation of top hit compounds 9998128 and 16040294 are shown listed in files 9998128.doc and 16040294.doc in SI.

For calculating van der Waals (vdW) forces the cutoff of 1.4 nm was adopted, while the particle-mesh Ewald (PME) summation method [72] was employed to compute the electrostatic energy.

The leapfrog algorithm [73] was employed to solve the corresponding Langevin equations. Every 10 fs we updated the pair-list for long-range interactions using the cutoff of 1.0 nm. To neutralize the system 12, 9, 8, 1, 6 and 5  $\text{Na}^+$  ions were, respectively, added to 2MXU, 1M4H, 4EMA, 4Y6K, 4K6I and 1H22.

After energy minimization with the steepest descent method [74], we performed position-restrained simulations for 500 ps allowing water molecules to get into the binding site. The Berendsen algorithm [75] and damping coefficient 0.1 ps were used to keep temperature 300 K constant during 500 ps NVT simulation. For production NPT runs at 300K and 1 atm the Parrinello-Rahman coupling [76] was employed with damping coefficient 0.5 ps.

### *Steered molecular dynamics (SMD)*

SMD is an useful tool to study single biomolecules using the external force as an additional variable [8, 9]. It can be also employed to probe the binding affinity by pulling ligand from the binding site of receptor [7]. It was recognized that the accuracy of SMD is compatible with that of the MM-PBSA method but its computational speed is much higher [10, 11, 13]. For the simulation setup in this paper SMD is about 133-fold faster than MM-PBSA (see a more detailed discussion in SI). Therefore, SMD can be used to refine docking results in virtual screening.

In SMD the ligand is attached to a dummy atom via a spring with spring constant  $k$  and the dummy atom is moved with a constant velocity  $v$  along the direction allowing a smooth exit from the binding site. Thus, during pulling the force, exerted by the dummy atom on the ligand, is  $F = k(\Delta x - vt)$ , where  $\Delta x$  is a displacement of pulled atom from the initial position.

Following AFM experiment [77] we chose  $k = 600 \text{ kJ}/(\text{mol} \cdot \text{nm}^2)$  and  $v = 5 \text{ nm/ns}$  which was used previously [10, 11, 13]. To prevent the receptor from drifting together with the ligand during pulling we restrained all its  $\text{C}\alpha$ -atoms but maintaining flexibility of side chain.



Possible pathways for ligand to escape from the binding pocket were determined using CAVER 3.0 [78], a plugin of Pymol . The easiest or optimum path with the lowest rupture force has been chosen [12]. It should be aware that in general the ligand changes direction during exit from the binding site. Therefor the drawback of SMD with a single pulling direction is that it does not take into account multi-directional movement . Motivated by this challenge Yang *et al.* [79] and Gu *et al.* [80] have proposed a SMD method with adaptive direction adjustments where the optimum path of ligand is navigated by minimizing the pulling force automatically during the simulation. For the cytochrome P450 3A4-metyrapone complex the multidirectional pulling provided the pathway with smaller rupture force than that predicted by the standard SMD [79]. The self-adaptive SMD also yielded a good correlation between the rupture forces and experimentally measured binding free energies for two sets of protein-ligand complexes [80]. Because the correlation level provided by the standard SMD [13] is compatible with that of self-adaptive SMD we have chosen SMD with a single pulling direction which is implemented in Gromacs.

Receptor-ligand complexes were solvated in boxes fulfilled by water. The box sizes of all targets are displayed in Table S1 in SI. After equilibration 500 ns SMD simulations were performed in the NPT mode. These runs are long enough to completely get the ligand out from the active site. To obtain good statistics five independent runs were carried out starting from the same initial conformation but with different velocity distributions.

In SMD one can choose either rupture force  $F_{\max}$  or non-equilibrium work  $W_{\text{pull}}$  as a scoring function to rank binding affinities. However, we will use the latter as it is more reliable [13].  $W_{\text{pull}}$  is defined as follows

$$W_{\text{pull}} = \int_0^{x_{\max}} \vec{F} \cdot d\vec{x} \approx \sum_1^{N_{\text{step}}} \frac{(F_{i+1} + F_i)}{2} (x_{i+1} - x_i), \quad (1)$$

where  $N_{\text{step}}$  is the total number of steps used in simulation. Thus instead of integral the summation by the trapezoidal rule is used for estimating the non-equilibrium work performed on the system.

#### *MM-PBSA method*

According this method [6] the binding free energy  $\Delta G_{\text{bind}}$  has the following terms (more details on this method may be found elsewhere [81, 82]):

$$\Delta G_{\text{bind}} = \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} + \Delta G_{\text{sur}} + \Delta G_{\text{PB}} - T\Delta S, \quad (2)$$

where  $\Delta E_{\text{elec}}$  and  $\Delta E_{\text{vdW}}$  are electrostatic and vdW interaction energies.  $\Delta G_{\text{PB}}$  and  $\Delta G_{\text{sur}}$  are polar and nonpolar solvation energies. The entropy change  $\Delta S$  was computed by the interaction entropy method proposed recently by Duan *et al.* [83].

#### *Definition of relevant quantities*

RMSD (root mean square deviation) is defined as the deviation of receptor backbone from its starting structure. HB (hydrogen bond) occurs if D (donor)-A (acceptor)  $A \leq 3.5 \text{ \AA}$ , H-A distance  $2.7 \leq \text{\AA}$  and D-H-A angle  $\geq 135$  degrees. If the distance between centers of mass of ligand and some receptor residue is within 0.65 nm then we assume that a non-bonded contact is formed.

## Results and Discussion

### *Novel protocol for efficient search in silico screening*

The major drawback of the SMD method is that it is capable to predict relative but not absolute binding free energies implying that based solely on its output one cannot justify about the strength of ligand binding. On the other hand, in SMD the larger is the non-equilibrium work  $W_{\text{pull}}$  the stronger is binding. Therefore, for a given target we propose to choose those compounds that have  $W_{\text{pull}}$  larger than that of a reference compound which has high binding affinity known from experiment. Only these prominent compounds will be further used for the multi-target design. This idea is illustrated as the last step highlighted in red in Figure 1. For six studied targets the reference compounds have the inhibition constant in the nM range (Table 1).

### *Simulation results*

#### *Docking results*

The total number of compounds from the PubChem [47] database is reduced to 5732 drug-like compounds after application of Lipinski's rule (Figure 1). Autodock Vina software has been used to dock ligands from the reduced set 5 targets involving 2MXU, 1M4H, 4EMA, 4K6I, and 4Y6K. The distributions of docking binding energy  $\Delta E_{\text{bind}}$  of 5327 ligands are shown in Figure S3 in SI. The binding energies to 2MXU vary from -1.4 to -11.9 kcal/mol, whereas  $-1.2 \leq \Delta E_{\text{bind}} \leq -11.3$  kcal/mol (1M4H),  $-1.3 \leq \Delta E_{\text{bind}} \leq -9.8$  kcal/mol (4Y6K),  $-1.4 \leq \Delta E_{\text{bind}} \leq -10.6$  kcal/mol (4EMA), and  $-1.0 \leq \Delta E_{\text{bind}} \leq -10.6$  kcal/mol (4K6I).

The binding energies of 5327 compounds to the A $\beta$  protofibril are correlated with those obtained for targets 1M4H (correlation level  $R=0.95$ ), 4EMA ( $R=0.80$ ) and 4Y6K ( $R=0.93$ ), while the poor correlation was seen for 4K6I ( $R=0.40$ ) (Fig. S4 in SI). Thus, as expected, binding affinities are sensitive to targets.

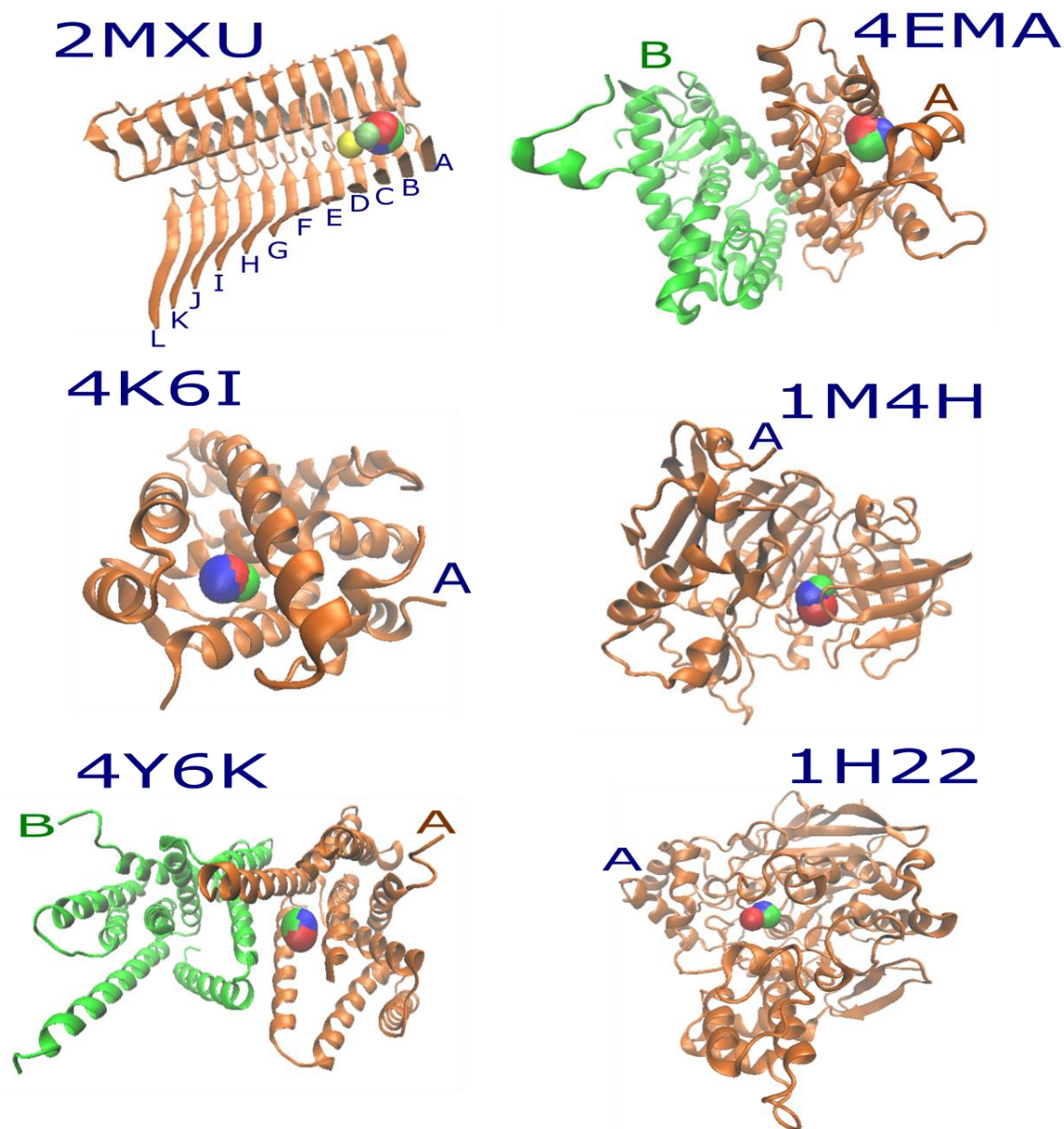


Figure 2. Position of the reference compound (red), CID 16040294 (blue), and CID 9998128 (green) in the best docking mode. For fibril 2MXU we show all ten top hits (see below). 4EMA and 4Y6K have two chains A (brown) and B (green). The A $\beta$ 11-42 fibril (2MXU) contains 12 peptides named by letters A-L. Except the A $\beta$  fibril case the binding site is known from experiment for receptors.

As seen below, compounds 16040294 and 998128 have been identified as the top leads, we show their position in the best docking mode of six targets together with the reference compounds (Figure 2). Because the A $\beta$  fibril does not have the well-defined binding site [20], ligands are located at different positions but preferring to stay next to strongly hydrophobic residues like Ile, Val and Leu (Figures S5, S6 and S7).

The binding energies, obtained in docking simulation, for six reference compounds to their targets vary between -7.2 and -12.6 kcal/mol (Table 1). The correlation between  $\Delta E_{\text{bind}}$  and experimental binding free energies is relatively low,  $R=-0.63$  (Fig. S8 in SI) implying that the docking method is not accurate enough due to crude approximations such as omission of receptor dynamics and limited number of trials for ligand position.

The hydrogen bond (HB) and non-bonded contact networks formed by 16040294, 9998128 and reference compounds with six targets are presented in Figures S5, S6 and S7, respectively. These figures have been prepared using LigPlot+ version 1.4.4 [84]. Compound 16040294 forms 2, 3, 0, 1, 2, and 1 HBs with 2MXU, 4EMA, 4K6I, 1M4H, 4Y6K, and 1H22, respectively, while for non-bonded contacts the corresponding numbers are 13, 9, 18, 12, 8, and 11 (Figure S5). Thus HB network is much poorer than the non-bonded contact network implying that the hydrogen bonding plays a less important role in stabilization of receptor-ligand complexes compared to non-bonded bonds. This is also valid for 9998128 (Figure S6) and reference compounds (Figure S7) which have a few or none HBs with targets. Compound 9998128 forms 3 HBs with 1M4H only, while just one HB occurs in complexes of reference compound with targets 4EMA, 4K6I, and 1M4H.

For 2MXU 16040294, 9998128 and reference compound Curcumin have 8 common non-bonded contacts with Val12(A), Val12(B), His14(B), Leu17(B), Leu17(D), Ile32 (B), Gly33(C), and Leu34(B) from peptides A, B, and C (Table S2 in SI). Compounds 16040294, 9998128 and the corresponding reference compound have 6, 7, 5, 7 and 9 common contacts with 4EMA, 4K6I, 1M4H, 4Y6K, and 1H22, respectively.

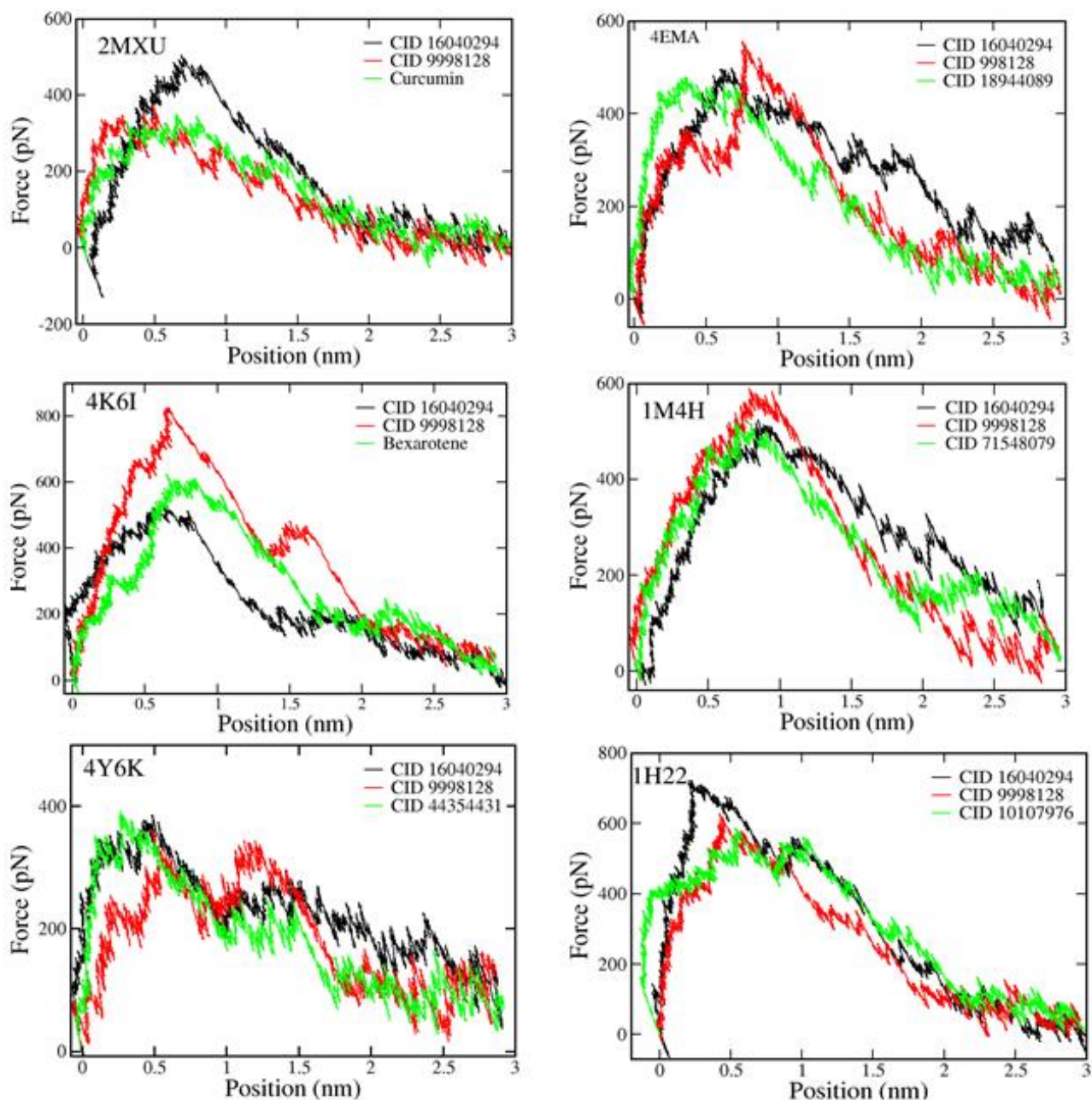
For further virtual screening we have chosen only those compounds, which have  $\Delta E_{\text{bind}} < -8.0$  kcal/mol, because the binding energy  $\approx -8$  kcal/mol corresponds to experimental  $\text{IC}_{50} \sim \mu\text{M}$ . Then we obtained 792, 834, 1417, 540, and 121 compounds for 2MXU, 4K4I, 1M4H, 4EMA, and 4Y6K, respectively (Figure 1). From these sets we found only 36 compounds which are common for all five targets and they have been selected for the next step of top hit prediction.

### *Blood brain barrier*

As AD drug candidates selected compounds should be able to travel to the brain or to overcome the BBB. The capability of crossing the BBB of the 36 common compounds was probed using the server PreADMET (see <https://preadmet.bmdrc.kr/adme/>). Keeping only those compounds, which have  $\log(\text{BB}) > 0$ , we have found 10 ligands (Figure 1) as top hits. Their ID and structures are shown in SI (Table S3).

### *SMD results*

Using Caver 3.0 software [78] we have found the optimal pulling direction for each complex. Some of them are shown in Figure S9 for targets 4K6I and 4Y6K, where the ligand is in the best docking mode with receptor. In order to obtain reliable result for  $W_{\text{pull}}$  five SMD simulations were carried out and the non-equilibrium work was averaged over all runs.



**Figure 4.** Typical force-position profiles for CID 9998128, CID 16040294 and the reference compounds (green curve) bound to multiple targets.

Shown in Figure 4 are typical force-position profiles for CID 9998128, CID 16040294 and the reference compounds bound to six targets. Clearly  $F_{\max}$  is sensible to both ligand and receptor. It also depends on SMD runs [12] (results not shown). Force-time/position curves were used to estimate the non-equilibrium work given by Eq. (1) for ranking binding affinities.

Table 3. Non-equilibrium works, obtained by SMD, for 10 compounds and the reference compound Curcumin (red) for target 2MXU. Results were averaged over 5 independent MD simulations. The binding free energy of Curcumin to A $\beta$  fibril is -13.3 kcal/mol [58].  $\Delta E_{\text{bind}}$  (kcal/mol) obtained by docking simulation is also shown. The ranking in docking simulation is in brackets in the last column.

No	Ligand	Work (kcal/mol)	$\Delta E_{\text{bind}}$ (kcal/mol) docking
1	<b>16040294</b>	<b>82.0 <math>\pm</math> 1.6</b>	<b>-8.3 (9)</b>
2	9549303	74.9 $\pm$ 3.7	-9.1 (4)
3	11494412	71.9 $\pm$ 3.8	-8.5 (8)
4	444746	68.2 $\pm$ 4.8	-8.6 (7)
5	11545419	63.1 $\pm$ 2.3	-8.7 (5)
6	<b>9998128</b>	<b>61.7 <math>\pm</math> 2.0</b>	<b>-9.7 (1)</b>
7	447767	60.5 $\pm$ 2.9	-9.5 (2)
8	<b>Curcumin</b>	<b>59.3 <math>\pm</math> 2.8</b>	-7.2 (11)
9	6419766	58.5 $\pm$ 1.7	-9.3 (3)
10	16122633	42.7 $\pm$ 1.9	-8.2 (10)
11	11790	36.0 $\pm$ 1.9	-8.6 (6)

For the A $\beta$  target (2MXU), the SMD method predicts that compound CID 16040294 is champion having the largest  $W_{\text{pull}}$  of 82 kcal/mol (Table 3), while the reference compound Curcumin is ranked 8<sup>th</sup>. Therefore, 7 compounds with  $W_{\text{pull}}$  larger than that of Curcumin were selected as top-hits for destroying A $\beta$  fibril. Note that the docking-based ranking of binding affinities is different

from that predicted by SMD (Table 3). The compound CID 9998128 is champion in docking but it runs seventh in SMD. Although CID 16040294 is the strongest in SMD, it is just ninth in docking.

For PPAR $\gamma$  (4EMA) target, we have chosen 5 compounds, which have  $W_{\text{pull}}$  larger than that of the reference compound CID 18944089 (Table S4 in SI), for the next screening step. Note that the inhibition constant of CID 18944089 is  $K_i = 0.4$  nM [61] corresponding to  $\Delta G_{\text{exp}} \approx -12.9$  kcal/mol. As in the A $\beta$  case SMD and docking simulation provide different rankings for binding affinities (Table S4).

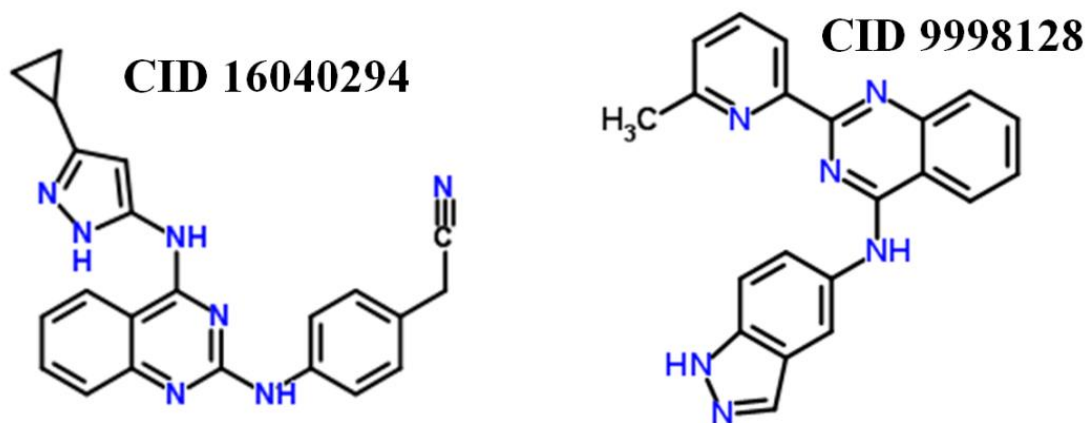
The reference compound Bexarotene was experimentally shown to strongly bind to RXR $\alpha$  (4K6I) target [62]. The present SMD simulation revealed that among 11 studied compounds Bexarotene is ranked third being worse than compounds 9998128 and 6419766 (Table S5 in SI). Therefore, for RXR $\alpha$  we kept only these two compounds for multi-target screening.

Similar to the RXR $\alpha$  case, for  $\beta$ -secretase (1M4H) we obtained two compounds that show stronger binding affinity than the reference compound 71548079 (Table S6). For  $\gamma$ -secretase (4Y6K) target, the reference compound 44354431 is at position 6 (Table S7) implying that 5 compounds should be kept for the final step of screening.

#### *Prediction of top leads*

As follows from Table 3 and Tables S4-S7 only compound 9998128 has  $W_{\text{pull}}$  exceeding that of reference compounds in all 5 targets. Compound 16040294 is better than the reference compound for all targets except RXR $\alpha$ , where it stands just after Bexarotene (Table S5). Therefore, together with 9998128 we will consider 16040294 as top lead for five targets. Their chemical structures (Figure 5) are more complex than Curcumin and Bexarotene (Figure S2) but less complex than reference compounds of targets PPAR $\gamma$ ,  $\beta$ - and  $\gamma$ -secretase. According to the standard strategy of designing multi-target leads, where structures of single-target leads are combined, structures of multi-target ligands should be more complicated compared to individual components. Overall, this is not consistent with our predictions based on searching drug candidates among big databases.





**Figure 5.** 2D chemical structures of top leads 9998128 and 16040294.

In comparison with 16040294, compound 9998128 displays a weaker binding affinity to 4EMA, 2MXU and 4Y6K but an opposite situation takes place in the 4K6I case where it has remarkably larger non-equilibrium work (Table S8 in SI). The SMD simulation predicts that 16040294 binds to 1M4H as strong as 9998128 because they have nearly the same pulling works.

Two top leads are as good as Bexarotene but much better than the rest of reference compounds in crossing BBB (Table S9 in SI). This prediction is based on the QSAR analysis and should be checked by experiment.

*The van der Waals (vdW) interaction drives binding of top leads towards multiple targets*

The time evolution of the vdW and electrostatic interaction energy between receptor and ligand during SMD simulation is depicted in Figure S10. Ligands remain in the binding pocket during the first  $\approx 200 - 300$  ps depending on complexes. Obviously, for this period the absolute value of the vdW interaction is higher than that of the electrostatic interaction and this is valid for other studied systems (results not shown). Therefore, the vdW interaction is more important than the electrostatic interaction in binding of two top leads and reference compounds towards six targets.

*High binding affinity of top hits to Acetylcholinesterase: SMD and MM-PBSA simulations*

Because AChE is a very popular target for designing AD drugs it is worthy to know whether hit compounds 9998128 and 16040294 are good binders to it. We first performed the SMD simulation to pull them and the reference compound 10107976 (Table 1) from the binding pocket of the PDB structure 1H22. Within error bars  $W_{\text{pull}}$  of 16040294 is compatible with the reference compound 10107976 (Table S10 in SI) suggesting that this lead strongly binds to AChE.

Because the nonequilibrium work of 9998128 is below that of the reference compound (Table S10), which has high binding affinity ( $K_i = 0.03$  nM, Table 1), it remains unclear about its



binding capability to 1H22. In order to clarify this issue we have computed  $\Delta G_{\text{bind}}$  using the MM-PBSA method. The 100 ns MD simulations were performed the force field AMBER-f99SB-ILDN [66] and TIP3P water model [67]. For each receptor-ligand complex four independent runs have been made starting from the same starting configuration (Figure 2), but with different seed numbers. To determine equilibration time  $t_{\text{eq}}$ , we considered the time dependence of C $\alpha$ -RMSD (Figure S11 in SI). Defining  $t_{\text{eq}}$  as a time when RMSD reaches saturation, we see that it varies between about 40 -60 ns depending on the system and MD runs. Snapshots collected at equilibrium (after  $t_{\text{eq}}$ ) have been used to estimate  $\Delta G_{\text{bind}}$  by Eq. (2).

The nonpolar ( $\Delta G_{\text{sur}}$ ) and polar ( $\Delta G_{\text{PB}}$ ) solvation energies are not sensitive to ligands, while the contribution of entropic term is larger for 16040294 than for 9998128 (Table S11 in SI). In accord with the SMD simulation (Figure S10) the electrostatic interaction is less important than the vdW interaction one because the absolute value of  $\Delta E_{\text{vdW}}$  is much higher than  $\Delta E_{\text{elec}}$ . Because  $\Delta G_{\text{bind}} < -20$  kcal/mol both top leads are expected to bind to AChE strongly.

## Conclusion

We have proposed a simple approach to rank ligand binding affinities by comparing non-equilibrium works with that of the reference compound whose inhibition constant was known from experiment. The main advantage of our trick is that the computation of  $W_{\text{pull}}$  is much less time consuming than  $\Delta G_{\text{bind}}$  and it can be used, therefore, as a power tool for computer-aided drug design for multi-target diseases. Having applied the multi-step screening procedure involving the new protocol at the last stage we have obtained two compounds 16040294 and 9998128 as potential drugs to modulate distinctive pathways leading to AD. We demonstrated that the vdW interaction is dominating over the electrostatic interaction in stabilizing their complex with all studied targets.

It is interesting to note that CID 9998128, mentioned in US2014031547 patent [85], inhibits activity casein kinase 1 delta (CK1  $\delta$ ) and may be used in treatment of neurodegenerative disorders such as Alzheimer's disease. Nitrile-containing CID 16040294 is inhibitor of Aurora-2 and GSK-3 and it is, therefore, a potential drug for cancer, diabetes and Alzheimer's disease [86].

We considered five targets related to molecular and genetic imbalances in AD patients and AChE target. The over-expression of AChE around A $\beta$  plaques may be associated with a disturbance in homeostasis of calcium [87] suggesting that AChE belongs to cellular imbalance. Thus, contrary to previous works, our study covers all categories of imbalances leading to AD.

Because the top leads have been identified by *in silico* experiments it would be important to probe their activity by *in vitro* as well as by *in vivo* experiments. However, our major goal is to show, as a proof of concept, an efficient way to seek potential drugs combining SMD with available experimental data on reference compounds.

## Competing Interest

The authors declare no competing financial interest.

## Acknowledgements

This work was supported by Department of Science and Technology at Ho Chi Minh city, Vietnam, and the Polish NCN grant 2015/19/B/ST4/02721, Poland. Allocation of CPU time at the supercomputer center TASK in Gdansk (Poland) is highly appreciated.

## Appendix A. Supplementary Information

Comparison of SMD efficiency with MM-PBSA is provided. The sizes boxes, used in SMD simulation are listed in Table S1. Table S2 provides the list of residues forming non-bonded contact with 16040294, 9998128 and reference compound for all targets. Table S3 presents the 3D structures of 10 ligands that have  $\log(\text{BB}) > 0$  and  $\Delta E_{\text{bind}} < -8.0$  kcal/mol to all 5 studied targets. Tables S4-S7 provide ranking of binding affinities in SMD and docking for target 4EMA (PPAR $\gamma$ ), 4K6I (RXR $\alpha$ ), 1M4H ( $\beta$ -secretase), and 4Y6K ( $\gamma$ -secretase), respectively. Nonequilibrium works to pull two drug candidates from six targets are shown in Table S8, while Table S9 provides  $\log(\text{BB})$  of two leads and six reference compounds. Table S10 gives nonequilibrium works to pull two leads and the reference compound 10107976 from the Acetylcholinesterase (1H22). The binding free energy, estimated by the MM-PBSA method, for two best candidates for target 1H22 is shown in Table S11. Figure S1 shows 3D PDB structures of six targets 2MXU, 4EMA, 4K6I, 1M4H, 4Y6K, and 1H22, while 2D structures of six reference compounds are presented in Figure S2. Figure S3 shows distributions of  $\Delta E_{\text{bind}}$  of 5732 ligands to five receptors provided by the docking method. Relationship between docking binding energies to 2MXU and other receptors is shown in Figure S4. Networks of non-bonded contacts and HBs are shown in Figures S5 and S6 for lead compounds 16040294 and 9998128, while Figure S7 for the reference compounds. Correlation between docking binding energies and experimental free energies is given in Figure S8. Figure S9 shows representative pulling directions. Time evolution of vdW and electrostatic interaction energies between receptor and ligand during SMD simulation for typical complexes is displayed in Figure S10. Figure S11 presents the time dependences of RMSD of 1H22 in complex with CID 16040294 and CID 9998128. Parameters used in MD simulation for compounds 16040294 and 9998128 are available in files 16040294.doc and 9998128.doc.

Supplementary data associated with this article can be found, in the online version, at XXX

## References

- [1] Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 2001, 46, 3-26.
- [2] Lee, F.S., Chu, Z.-T., Bolger, M.B., Warshel, A. Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng.* 1992, 5, 215-28.
- [3] Åqvist, J., Medina, C., Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 1994, 7, 385-91.
- [4] Kirkwood, J.G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 1935, 3, 300-13.
- [5] Zwanzig, R.W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* 1954, 22, 1420-6.
- [6] Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 2000, 33, 889-97.
- [7] Grubmüller, H., Heymann, B., Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science.* 1996, 997-9.
- [8] Isralewitz, B., Gao, M., Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 2001, 11, 224-30.
- [9] Kumar, S., Li, M.S. Biomolecules under mechanical force. *Phys. Rep.* 2010, 486, 1-74.
- [10] Mai, B.K., Li, M.S. Neuraminidase inhibitor R-125489—a promising drug for treating influenza virus: steered molecular dynamics approach. *Biochem. Biophys. Res. Commun.* 2011, 410, 688-91.
- [11] Suan Li, M., Khanh Mai, B. Steered molecular dynamics-a promising tool for drug design. *Curr. Bioinform.* 2012, 7, 342-51.
- [12] Mai, B.K., Viet, M.H., Li, M.S. Top leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach. *J. Chem. Inf. Model.* 2010, 50, 2236-47.
- [13] Vuong, Q.V., Nguyen, T.T., Li, M.S. A new method for navigating optimal direction for pulling ligand from binding pocket: application to ranking binding affinity by steered molecular dynamics. *J. Chem. Inf. Model.* 2015, 55, 2731-8.
- [14] Katzman, R. The prevalence and malignancy of Alzheimer disease: a major killer. *Arch. Neurol.* 1976, 33, 217-8.
- [15] Hardy, J., Selkoe, D.J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *science.* 2002, 297, 353-6.
- [16] Herrup, K. The case for rejecting the amyloid cascade hypothesis. *Nat. Neurosci.* 2015, 794-9.
- [17] Francis, P.T., Palmer, A.M., Snape, M., Wilcock, G.K. The cholinergic hypothesis of Alzheimer's disease: a review of progress. *J. Neurol. Neurosurg. Psychiatry.* 1999, 66, 137-47.
- [18] Doig, A.J., Derreumaux, P. Inhibition of protein aggregation and amyloid formation by small molecules. *Curr. Opin. Struct. Biol.* 2015, 30, 50-6.
- [19] Viet, M.H., Ngo, S.T., Lam, N.S., Li, M.S. Inhibition of aggregation of amyloid peptides by beta-sheet breaker peptides and their binding affinity. *J. Phys. Chem. B.* 2011, 115, 7433-46.
- [20] Viet, M.H., Chen, C.-Y., Hu, C.-K., Chen, Y.-R., Li, M.S. Discovery of dihydrochalcone as potential lead for Alzheimer's disease: in silico and in vitro study. *PLoS ONE.* 2013, 8, e79151.

- [21] Bernstein, S.L., Dupuis, N.F., Lazo, N.D., Wytttenbach, T., Condrón, M.M., Bitan, G., et al. Amyloid- $\beta$  protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* 2009, 1, 326-31.
- [22] Walsh, D.M., Klyubin, I., Fadeeva, J.V., Cullen, W.K., Anwyl, R., Wolfe, M.S., et al. Naturally secreted oligomers of amyloid  $\beta$  protein potently inhibit hippocampal long-term potentiation in vivo. *Nature*. 2002, 416, 535-9.
- [23] Bursavich, M.G., Harrison, B.A., Blain, J.-F. Gamma secretase modulators: new Alzheimer's drugs on the horizon? *J. Med. Chem.* 2016, 59, 7389-409.
- [24] Ghosh, A.K., Osswald, H.L. BACE1 ( $\beta$ -secretase) inhibitors for the treatment of Alzheimer's disease. *Chem. Soc. Rev.* 2014, 43, 6765-813.
- [25] Golde, T.E., Koo, E.H., Felsenstein, K.M., Osborne, B.A., Miele, L.  $\gamma$ -Secretase inhibitors and modulators. *BBA-BIOMEMBRANES*. 2013, 1828, 2898-907.
- [26] Schellenberg, G.D., Montine, T.J. The genetics and neuropathology of Alzheimer's disease. *Acta Neuropathol.* 2012, 124, 305-23.
- [27] Tanzi, R.E. The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med.* 2012, 2, a006296.
- [28] Michalik, L., Auwerx, J., Berger, J.P., Chatterjee, V.K., Glass, C.K., Gonzalez, F.J., et al. International Union of Pharmacology. LXI. Peroxisome proliferator-activated receptors. *Pharmacol. Rev.* 2006, 58, 726-41.
- [29] Alagiakrishnan, K., Gill, S.S., Fagarasanu, A. Genetics and epigenetics of Alzheimer's disease. *Postgrad. Med. J.* 2012, 88, 522-9.
- [30] Kitamura, Y., Shimohama, S., Koike, H., Kakimura, J.-i., Matsuoka, Y., Nomura, Y., et al. Increased expression of cyclooxygenases and peroxisome proliferator-activated receptor- $\gamma$  in Alzheimer's disease brains. *Biochem. Biophys. Res. Commun.* 1999, 254, 582-6.
- [31] Jiang, Q., Heneka, M., Landreth, G.E. The role of peroxisome proliferator-activated receptor- $\gamma$  (PPAR $\gamma$ ) in Alzheimer's disease. *CNS drugs*. 2008, 22, 1-14.
- [32] Yan, Q., Zhang, J., Liu, H., Babu-Khan, S., Vassar, R., Biere, A.L., et al. Anti-inflammatory drug therapy alters  $\beta$ -amyloid processing and deposition in an animal model of Alzheimer's disease. *J. Neurosci.* 2003, 23, 7504-9.
- [33] Heneka, M.T., Sastre, M., Dumitrescu-Ozimek, L., Hanke, A., Dewachter, I., Kuiperi, C., et al. Acute treatment with the PPAR $\gamma$  agonist pioglitazone and ibuprofen reduces glial inflammation and A $\beta$ 1-42 levels in APPV717I transgenic mice. *Brain*. 2005, 128, 1442-53.
- [34] Mandrekar-Colucci, S., Landreth, G.E. Nuclear receptors as therapeutic targets for Alzheimer's disease. *Expert Opin. Ther. Targets*. 2011, 15, 1085-97.
- [35] Koldamova, R.P., Lefterov, I.M., Staufenbiel, M., Wolfe, D., Huang, S., Glorioso, J.C., et al. The Liver X Receptor Ligand T0901317 Decreases Amyloid  $\beta$  Production in Vitro and in a Mouse Model of Alzheimer's Disease. *Journal of Biological Chemistry*. 2005, 280, 4079-88.
- [36] Zheng, H., Fridkin, M., Youdim, M. New approaches to treating Alzheimer's disease. *Perspect Medicin Chem.* 2015, 7, 1.
- [37] Cavalli, A., Bolognesi, M.L., Minarini, A., Rosini, M., Tumiatti, V., Recanatini, M., et al. Multi-target-directed ligands to combat neurodegenerative diseases. *J. Med. Chem.* 2008, 51, 347-72.
- [38] Viayna, E., Gómez, T., Galdeano, C., Ramírez, L., Ratia, M., Badia, A., et al. Novel Huprine Derivatives with Inhibitory Activity toward  $\beta$ -Amyloid Aggregation and Formation as Disease-Modifying Anti-Alzheimer Drug Candidates. *ChemMedChem*. 2010, 5, 1855-70.

- [39] Bolognesi, M.L., Cavalli, A., Valgimigli, L., Bartolini, M., Rosini, M., Andrisano, V., et al. Multi-target-directed drug design strategy: from a dual binding site acetylcholinesterase inhibitor to a trifunctional compound against Alzheimer's disease. *J. Med. Chem.* 2007, 50, 6446-9.
- [40] Bolognesi, M.L., Bartolini, M., Tarozzi, A., Morroni, F., Lizzi, F., Milelli, A., et al. Multitargeted drugs discovery: balancing anti-amyloid and anticholinesterase capacity in a single chemical entity. *Bioorg. Med. Chem. Lett.* 2011, 21, 2655-8.
- [41] Rabal, O., Sánchez-Arias, J.A., Cuadrado-Tejedor, M., De Miguel, I., Pérez-González, M., Garcia-Barroso, C., et al. Design, synthesis, and biological evaluation of first-in-class dual acting histone deacetylases (HDACs) and phosphodiesterase 5 (PDE5) inhibitors for the treatment of Alzheimer's disease. *J. Med. Chem.* 2016, 59, 8967-9004.
- [42] Domínguez, J.L., Fernández-Nieto, F., Castro, M., Catto, M., Paleo, M.R., Porto, S., et al. Computer-Aided Structure-Based Design of Multitarget Leads for Alzheimer's Disease. *J. Chem. Inf. Model.* 2014, 55, 135-48.
- [43] Prati, F., Cavalli, A., Bolognesi, M.L. Navigating the chemical space of multitarget-directed ligands: From hybrids to fragments in Alzheimer's disease. *Molecules.* 2016, 21, 466.
- [44] Morphy, R., Kay, C., Rankovic, Z. From magic bullets to designed multiple ligands. *Drug discovery today.* 2004, 9, 641-51.
- [45] Hann, M.M., Leach, A.R., Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of chemical information and computer sciences.* 2001, 41, 856-64.
- [46] Bottegoni, G., Favia, A.D., Recanatini, M., Cavalli, A. The role of fragment-based and computational methods in polypharmacology. *Drug discovery today.* 2012, 17, 23-34.
- [47] Bolton, E.E., Wang, Y., Thiessen, P.A., Bryant, S.H. PubChem: integrated platform of small molecules and biological activities. *Annu Rep Comput Chem.* 2008, 4, 217-41.
- [48] Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 2012, 64, 4-17.
- [49] Nasica-Labouze, J., Nguyen, P.H., Sterpone, F., Berthoumieu, O., Buchete, N.-V., Coté, S.b., et al. Amyloid  $\beta$  protein and Alzheimer's disease: When computer simulations complement experimental studies. *Chem. Rev.* 2015, 115, 3518-63.
- [50] Xiao, Y., Ma, B., McElheny, D., Parthasarathy, S., Long, F., Hoshi, M., et al. A  $[\beta](1-42)$  fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* 2015, 22, 499-505.
- [51] Tycko, R. Alzheimer's disease: Structure of aggregates revealed. *Nature.* 2016.
- [52] Liberato, M.V., Nascimento, A.S., Ayers, S.D., Lin, J.Z., Cvaro, A., Silveira, R.L., et al. Medium chain fatty acids are selective peroxisome proliferator activated receptor (PPAR)  $\gamma$  activators and pan-PPAR partial agonists. *PLoS One.* 2012, 7, e36297.
- [53] Boerma, L.J., Xia, G., Qui, C., Cox, B.D., Chalmers, M.J., Smith, C.D., et al. Defining the communication between agonist and coactivator binding in the retinoid X receptor  $\alpha$  ligand binding domain. *J. Biol. Chem.* 2014, 289, 814-26.
- [54] Bai, X.-c., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., et al. An atomic structure of human [ggr]-secretase. *Nature.* 2015.
- [55] Dang, S., Wu, S., Wang, J., Li, H., Huang, M., He, W., et al. Cleavage of amyloid precursor protein by an archaeal presenilin homologue PSH. *Proc. Natl. Acad. Sci. USA.* 2015, 112, 3344-9.

- [56] Li, X., Dang, S., Yan, C., Gong, X., Wang, J., Shi, Y. Structure of a presenilin family intramembrane aspartate protease. *Nature*. 2013, 493, 56-61.
- [57] Hong, L., Turner, R.T., Koelsch, G., Shin, D., Ghosh, A.K., Tang, J. Crystal structure of memapsin 2 ( $\beta$ -secretase) in complex with an inhibitor OM00-3. *Biochemistry*. 2002, 41, 10963-7.
- [58] Yang, F., Lim, G.P., Begum, A.N., Ubeda, O.J., Simmons, M.R., Ambegaokar, S.S., et al. Curcumin inhibits formation of amyloid  $\beta$  oligomers and fibrils, binds plaques, and reduces amyloid in vivo. *J. Biol. Chem.* 2005, 280, 5892-901.
- [59] Abdel-Magid, A.  $\beta$ -Secretase Inhibitors for the Treatment of Alzheimer's Disease and Down's Syndrome. *ACS Med. Chem. Lett.* 2013, 4, 578-9.
- [60] Owens, A.P., Nadin, A., Talbot, A.C., Clarke, E.E., Harrison, T., Lewis, H.D., et al. High affinity, bioavailable 3-amino-1, 4-benzodiazepine-based  $\gamma$ -secretase inhibitors. *Bioorg. Med. Chem. Lett.* 2003, 13, 4143-5.
- [61] Buckle, D., Cantello, B., Cawthorne, M., Coyle, P., Dean, D., Faller, A., et al. Non thiazolidinedione antihyperglycaemic agents. 1:  $\alpha$ -Heteroatom substituted  $\beta$ -phenylpropanoic acids. *Bioorg. Med. Chem. Lett.* 1996, 6, 2121-6.
- [62] Farmer, L.J., Zhi, L., Jeong, S., Kallel, E.A., Croston, G., Flatten, K.S., et al. Synthesis and structure-activity relationships of potent conformationally restricted retinoid X receptor ligands. *Bioorg. Med. Chem. Lett.* 1997, 7, 2747-52.
- [63] Camps, P., Gómez, E., Muñoz-Torrero, D., Badia, A., Clos, M.V., Curutchet, C., et al. Binding of 13-amidohuprines to acetylcholinesterase: Exploring the ligand-Induced conformational change of the Gly117-Gly118 peptide bond in the oxyanion hole. *J. Med. Chem.* 2006, 49, 6833-40.
- [64] Sanner, M.F. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* 1999, 17, 57-61.
- [65] 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.* 2010, 31, 455-61.
- [66] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010, 78, 1950-8.
- [67] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983, 79, 926-35.
- [68] Wang, J., Wang, W., Kollman, P.A., Case, D.A. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* 2001, 222, U403.
- [69] da Silva, A.W.S., Vranken, W.F. ACPYPE-Antechamber python parser interface. *BMC Res Notes*. 2012, 5, 367.
- [70] Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* 2004, 25, 1157-74.
- [71] Jakalian, A., Bush, B.L., Jack, D.B., Bayly, C.I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 2000, 21, 132-46.
- [72] Darden, T., York, D., Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 1993, 98, 10089-92.
- [73] Hockney, R., Goel, S., Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* 1974, 14, 148-58.

- [74] Bartholomew-Biggs, M. The steepest descent method. *Nonlinear Optimization with Financial Applications*. 2005, 51-64.
- [75] Berendsen, H.J., Postma, J.v., van Gunsteren, W.F., DiNola, A., Haak, J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 1984, 81, 3684-90.
- [76] Parrinello, M., Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 1981, 52, 7182-90.
- [77] Gibson, C.T., Carnally, S., Roberts, C.J. Attachment of carbon nanotubes to atomic force microscope probes. *Ultramicroscopy*. 2007, 107, 1118-22.
- [78] Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* 2012, 8, e1002708.
- [79] Yang, K., Liu, X.L., Wang, X.C., Jiang, H.L. A steered molecular dynamics method with adaptive direction on adjustments. *Biochemical and Biophysical Research Communications*. 2009, 379, 494-8.
- [80] Gu, J., Li, H., Wang, X. A Self-Adaptive Steered Molecular Dynamics Method Based on Minimization of Stretching Force Reveals the Binding Affinity of Protein-Ligand Complexes. *Molecules*. 2015, 20, 19236-51.
- [81] Nguyen, T.T., Mai, B.K., Li, M.S. Study of Tamiflu sensitivity to variants of A/H5N1 virus using different force fields. *J. Chem. Inf. Model.* 2011, 51, 2266-76.
- [82] Ngo, S.T., Li, M.S. Curcumin binds to A $\beta$ 1–40 peptides and fibrils stronger than ibuprofen and naproxen. *J. Phys. Chem. B*. 2012, 116, 10165-75.
- [83] Duan, L., Liu, X., Zhang, J.Z.H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy. *Journal of the American Chemical Society*. 2016, 138, 5722-8.
- [84] Laskowski, R., Swindells, M. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* 2011, 51, 2778-86.
- [85] Sheridan, J.M., Heal, J.R., Hamilton, W.D., Pike, I. Casein kinase 1delta (ck 1delta) inhibitors and their use in the treatment of neurode-generative diseases such as tauopathies. *Electrophoretics Limited, US*, 2011, Vol. US2014031547 A1.
- [86] Fleming, F.F., Yao, L.H., Ravikumar, P.C., Funk, L., Shook, B.C. Nitrile-Containing Pharmaceuticals: Efficacious Roles of the Nitrile Pharmacophore. *Journal of Medicinal Chemistry*. 2010, 53, 7902-17.
- [87] Sberna, G., Sáez-Valero, J., Beyreuther, K., Masters, C.L., Small, D.H. The Amyloid  $\beta$ -Protein of Alzheimer's Disease Increases Acetylcholinesterase Expression by Increasing Intracellular Calcium in Embryonal Carcinoma P19 Cells. *J. Neurochem.* 1997, 69, 1177-84.

# **Protocol for fast screening of multi-target drug candidates: Application to Alzheimer's disease**

Nguyen Quoc Thai<sup>1,2,3</sup>, Hoang Linh Nguyen<sup>1</sup>, Huynh Quang Linh<sup>3</sup>, and Mai Suan Li<sup>1,4</sup>

*<sup>1</sup>Institute for Computational Sciences and Technology, SBI building, Quang Trung Software City, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam*

*<sup>2</sup>Dong Thap University, 783 Pham Huu Lau Street, Ward 6, Cao Lanh City, Dong Thap, Vietnam*

*<sup>3</sup>Biomedical Engineering Department, University of Technology -VNU HCM, 268 Ly Thuong Kiet Str., Distr. 10, Ho Chi Minh City, Vietnam*

*<sup>4</sup>Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*

## **SUPPORTING INFORMATION**



## Comparison of SMD efficiency with MM-PBSA

In SMD most of time is spent on SMD runs because calculation of the non-equilibrium work takes very little time. In MM-PBSA, beside generating MD trajectories the additional time, needed for estimating the binding free energy, depends on the way we compute the entropy term. If  $\Delta S$  is estimated by the expensive normal mode method then the computer time depends on system size and the number of selected snapshots but typically this time is comparable with that of MD run. In this work we used the method developed by Duan *et al.* (JACS 2016, 138, 5722–5728) where the entropy term is calculated directly from MD simulation without any extra computational cost. Therefore with this new method most of time is spent on MD runs. Here we have carried out 5 runs of 600 ps SMD simulation (3 ns total) and 4 trajectories of 100 ns MD simulation (400 ns total) and consequently SMD is  $400/3 \approx 133$  times faster than MM-PBSA. Note that MD and SMD simulations basically consume the same amount of computer time for a given time interval.

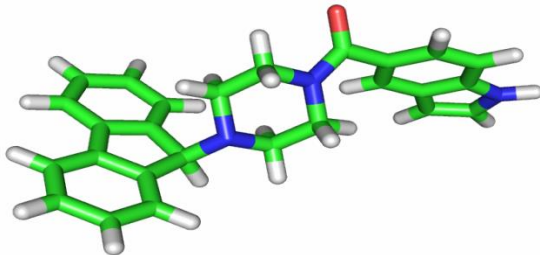
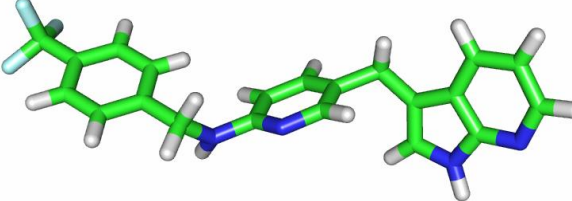
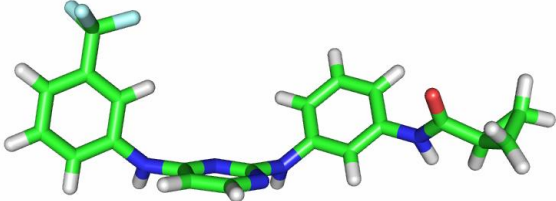
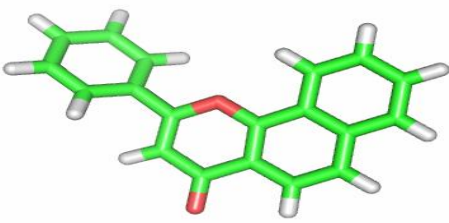
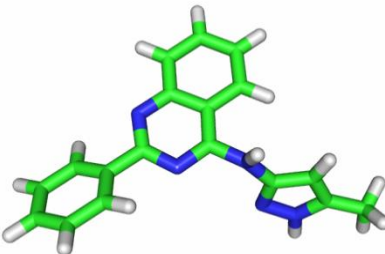
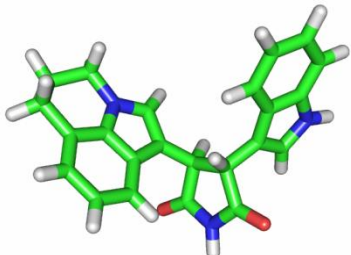
Table S1. The sizes boxes, used in SMD simulation, for all targets.

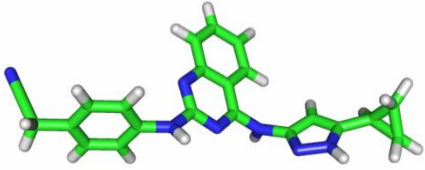
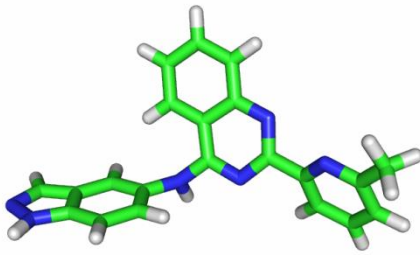
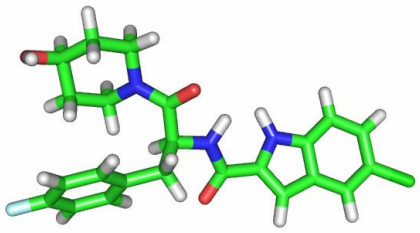
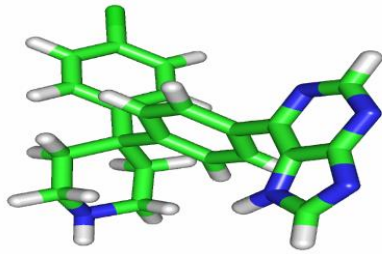
Target	Box size (nm <sup>3</sup> )
A $\beta$ protofibril (2MXU)	10.1x8.4x18.5
$\beta$ -secretase (1M4H)	4.6x8.4x15.4
$\gamma$ -secretase (4Y6K)	7.1x7.2x11.6
PPAR $\gamma$ (4EMA)	7.7x6.7x11.1
RXR $\alpha$ (4K6I)	6.7x6.8x10.7
AChE (1H22)	8.3x7.5x12.1

**Table S2.** List of residues forming non-bonded contact with 16040294, 9998128 and reference compound (blue). For a given target three ligands have common residues (red). The number of common residues is shown in brackets next to the target name in the first column.

2MXU (8)	16040294	Val12(A), Val12(B), His14(A), His14(B), Leu17(A), Leu17(B), Leu17(D), Ile32 (B), Ile32(C), Ile32(D), Gly33(C), Gly33(D), Leu34(B)
	9998128	Val12(A), Val12(B), His14(A), His14(B), His14(C), Leu17(B), Leu17(C), Leu17(D), Ile32(A), Ile32(B), Ile32(C), Gly33(A), Gly33(B), Gly33(C), Leu34(B)
	Curcumin	Val12(A), Val12(B), His14(B), Leu17(B), Leu17(D), Ile32(A), Ile32(B), Ile32(D), Gly33(A), Gly33(C), Gly33(D), Leu34(A), Leu34(B), Leu34(C), Leu34(D)
4EMA (6)	16040294	Ile281, Phe287, Arg288, Ser289, Tyr327, Leu330, Ile341, Met348, Met364
	9998128	Gly284, Cys285, Arg288, Ser289, Ile326, Tyr327, Leu330, Val339, Ile341, Ser342, Met348, Leu353, Met364
	18944089	Ile281, Phe282, Gly284, Cys285, Gln286, Arg288, Tyr327, Leu 330, Ile341, Ser342, Met348, Phe363, Met364, Lys367, His449, Leu469, Tyr473
4K6I (7)	16040294	Pro264, Val265, Ile268, Ala272, Gln275, Leu309, Ile310, Ser312, Phe313, Arg316, Val332, Ala337, Ala340, Val342, Ile345, Val349, Ile428, Cys432,
	9998128	Ile268, Cys269, Ala271, Ala272, Gln275, Trp305, Asn306, Leu309, Ile310, Phe313, Arg316, Leu326, Ala327, Val342, Ile345, Cys432, Leu436, Phe439
	Bexarotene	Ile268, Ala271, Ala272, Gln275, Leu309, Phe313, Arg316, Leu326, Ile345, Val349, Cys432, Leu436, His435, Phe439
1M4H (5)	16040294	Asp32, Gly34, Pro70, Tyr71, Thr72, Gln73, Phe108, Ile118, Tyr198, Asp228, Gly230, Asn233,
	9998128	Pro70, Tyr71, Thr72, Gln73, Phe108, Ile110, Trp115, Gly230, Arg235
	71548079	Gly34, Pro70, Tyr71, Thr72, Gln73, Lys107, Ile110, Asp228, Gly230, Thr231
4Y6K (7)	16040294	Phe57, Leu158, Ala159, Asp162, Val215, Met216, Met223, Leu275
	9998128	Leu54, Leu55, Phe57, Thr58, Leu158, Ala159, Tyr161, Asp162, Val215, Met216, Met223, Leu275
	44354431	Phe57, Phe79, Leu158, Ala159, Tyr161, Val215, Met216, Gly219, Met223, Leu275
1H22 (9)	16040294	Tyr70, Asp72, Trp84, Gly117, Gly118, Tyr121, Gly123, Tyr130, Phe330, Phe331, Tyr334
	9998128	Tyr70, Asp72, Trp84, Asn85, Gly117, Gly118, Tyr121, Ser122, Tyr130, Glu199, Phe330, Phe331, Tyr334, His440
	10107976	Tyr70, Asp72, Trp84, Gly118, Tyr121, Ser122, Tyr130, Glu199, Tyr279, Phe290, Phe330, Phe331, Tyr334

**Table S3.** 3D structures of 10 ligands that have  $\log(\text{BB}) > 0$  and  $\Delta E_{\text{bind}} < -8.0$  kcal/mol to all 5 targets. Two compounds in blue are top leads revealed by our new protocol.

No.	ID	logBB	3D structure
1	447767	0.89	
2	11545419	0.84	
3	9549303	0.53	
4	11790	0.46	
5	6419766	0.45	
6	11494412	0.40	

7	<b>16040294</b>	0.33	
8	<b>9998128</b>	0.25	
9	444746	0.17	
10	16122633	0.07	

**Table S4.** Ranking of binding affinities in SMD and docking for target 4EMA (PPAR $\gamma$ ). The reference compound CID 18944089 (red) has the experimental binding free energy  $\Delta G_{\text{exp}} = -12.9$  kcal/mol.

No.	Ligand	Work (kcal/mol)	$\Delta E_{\text{bind}}$ (kcal/mol) (ranking in docking)
1	9549303	$130.5 \pm 4.0$	-9.3(2)
2	<b>16040294</b>	<b><math>118.5 \pm 1.9</math></b>	<b>-8.1(11)</b>
3	444746	$115.6 \pm 9.2$	-8.6(5)
4	<b>9998128</b>	<b><math>99.3 \pm 3.4</math></b>	<b>-9.0(3)</b>
5	11545419	$96.9 \pm 4.8$	-9.6(1)
6	<b>18944089</b>	<b><math>92.6 \pm 4.5</math></b>	<b>-8.3(8)</b>
7	11494412	$86.1 \pm 3.9$	-8.9(4)
8	447767	$82.3 \pm 6.2$	-8.5(6)
9	6419766	$77.4 \pm 4.3$	-8.2(10)
10	16122633	$60.9 \pm 5.2$	-8.3(7)
11	11790	$43.1 \pm 1.8$	-8.2(9)

**Table S5.** The same as in Table S1 but for target 4K6I (RXR $\alpha$ ). The reference compound Bexarotene (red) has  $\Delta G_{\text{exp}} = -10.8$  kcal/mol.

No.	Ligand	Work (kcal/mol)	$\Delta E_{\text{bind}}$ (kcal/mol) (ranking in docking)
1	<b>9998128</b>	<b><math>132.3 \pm 3.4</math></b>	<b>-8.2(10)</b>
2	6419766	$120.7 \pm 2.8$	-8.2(8)
3	<b>Bexarotene</b>	<b><math>119.2 \pm 6.9</math></b>	-12.6(1)
4	<b>16040294</b>	<b><math>102.3 \pm 3.2</math></b>	<b>-8.2(9)</b>
5	16122633	$97.5 \pm 3.4$	-8.0(11)
6	11494412	$96.7 \pm 5.3$	-9.7(2)
7	9549303	$83.6 \pm 4.4$	-9.5(3)
8	447767	$83.0 \pm 2.2$	-8.2(7)
9	11545419	$81.2 \pm 4.2$	-9.4(4)
10	444746	$69.6 \pm 3.4$	-8.8(6)
11	11790	$67.4 \pm 2.2$	-9.3(5)

**Table S6.** The same as in Table S1 but for target 1M4H ( $\beta$ -secretase). The reference compound CID 71548079 (red) has  $\Delta G_{\text{exp}} = -14.4$  kcal/mol.

No.	Ligand	Work (kcal/mol)	$\Delta E_{\text{bind}}$ (kcal/mol) (ranking in docking)
1	16040294	$122.5 \pm 4.2$	-8.5(11)
2	9998128	$112.7 \pm 7.2$	-9.8(2)
3	71548079	$110.8 \pm 3.1$	-8.9(9)
4	11545419	$99.3 \pm 5.7$	-9.6(4)
5	447767	$98.4 \pm 7.1$	-10.5(1)
6	6419766	$79.6 \pm 5.2$	-9.4(6)
7	16122633	$73.5 \pm 4.4$	-9.0(8)
8	444746	$62.0 \pm 3.4$	-9.7(3)
9	11494412	$46.1 \pm 2.0$	-9.4(7)
10	9549303	$32.6 \pm 3.3$	-9.6(5)
11	11790	$32.1 \pm 2.5$	-8.8(10)



**Table S7.** The same as in Table S1 but for target 4Y6K ( $\gamma$ -secretase). The reference compound CID 44354431 (red) has  $\Delta G_{\text{exp}} = -12.7$  kcal/mol.

No.	Ligand	Work (kcal/mol)	$\Delta E_{\text{bind}}$ (kcal/mol) (ranking in docking)
1	16040294	$89.8 \pm 1.7$	-8.1(10)
2	9998128	$85.8 \pm 1.5$	-8.3(5)
3	444746	$80.0 \pm 4.7$	-8.3(6)
4	9549303	$79.8 \pm 4.0$	-8.5(2)
5	16122633	$77.6 \pm 3.2$	-8.3(7)
6	44354431	$73.5 \pm 5.0$	-8.5(1)
7	11494412	$64.6 \pm 1.8$	-8.5(3)
8	11545419	$63.6 \pm 3.7$	-8.4(4)
9	447767	$62.7 \pm 4.5$	-8.2(8)
10	6419766	$61.6 \pm 4.3$	-8.1(9)
11	11790	$50.4 \pm 2.3$	-8.0(11)

**Table S8.** Non-equilibrium works to pull two drug candidates from six targets.

Leads	Work (kcal/mol)					
	4EMA	2MXU	1M4H	4K6I	4Y6K	1H22
16040294	108.7	82.0	114.5	102.3	87.8	134.5
9998128	91.7	61.7	112.7	147.3	77.3	110.3

**Table S9.** Log(BB) of two leads (blue) and six reference compounds.

Compound	Target	Log(BB)	Molecular Weight (g/mol)
Curcumin	A $\beta$ 11-42	-1.04	368.38
Bexarotene	RXR $\alpha$	0.27	348.48
CID 18944089	PPAR $\gamma$	-0.13	432.47
CID 71548079	$\beta$ -secretase	-1.09	415.51
CID 44354431	$\gamma$ - secretase	-1.29	490.51
CID 10107976	Acetylcholinesterase	0.12	284.78
CID 16040294	All six targets	0.33	381.44
CID 9998128	All six targets	0.25	352.40

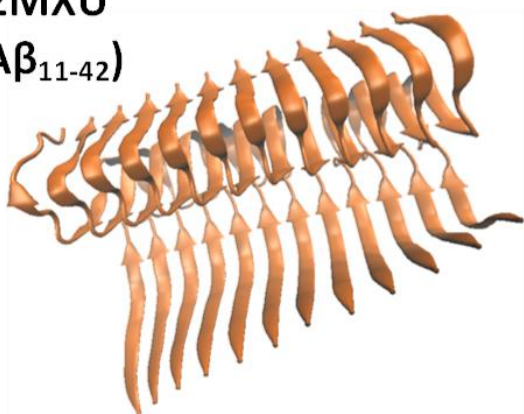
**Table S10.** Nonequilibrium works to pull two leads (blue) and the reference compound 10107976 (red) from the Acetylcholinesterase (1H22).

Ligand	Work (kcal/mol)
10107976	136.5 $\pm$ 3.5
16040294	134.5 $\pm$ 6.4
9998128	110.3 $\pm$ 3.3

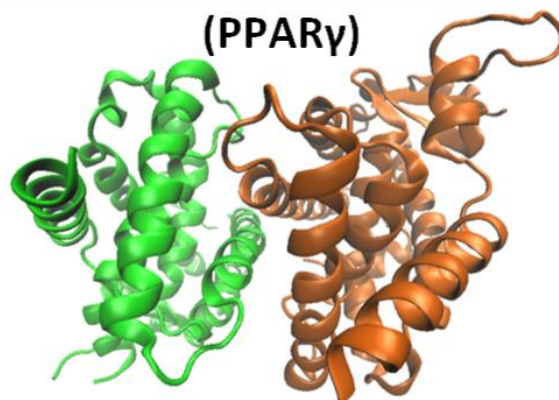
**Table S11.** Binding free energy  $\Delta G_{\text{bind}}$  (kcal/mol), estimated by the MM-PBSA method, for two best candidates for target 1H22.

Compounds CID	$\Delta E_{\text{elec}}$ (kcal/mol)	$\Delta E_{\text{vdW}}$ (kcal/mol)	$\Delta G_{\text{PB}}$ (kcal/mol)	$\Delta G_{\text{sur}}$ (kcal/mol)	$-T\Delta S$ (kcal/mol)	$\Delta G_{\text{bind}}$ (kcal/mol)
16040294	-16.43	-50.64	29.02	-6.55	18.04	-26.57
9998128	-13.20	-44.18	28.59	-6.41	12.91	-22.30

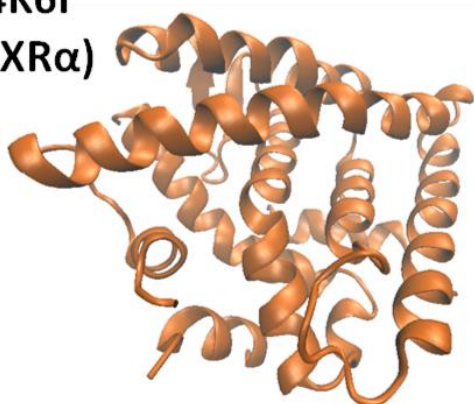
**2MXU**  
**(A $\beta$ <sub>11-42</sub>)**



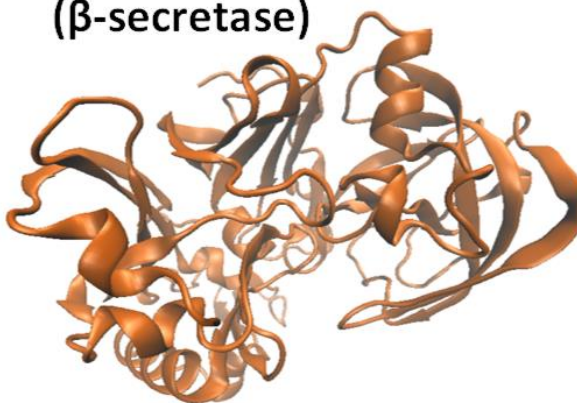
**4EMA**  
**(PPAR $\gamma$ )**



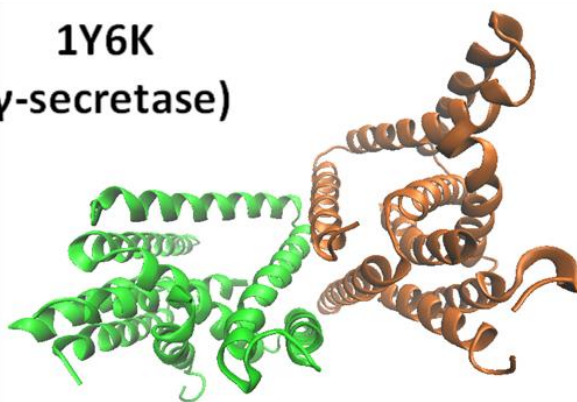
**4K6I**  
**(RXR $\alpha$ )**



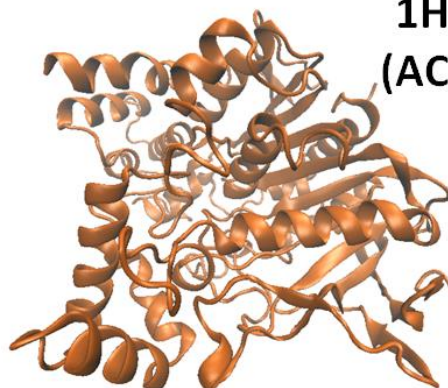
**1M4H**  
**( $\beta$ -secretase)**



**1Y6K**  
**( $\gamma$ -secretase)**

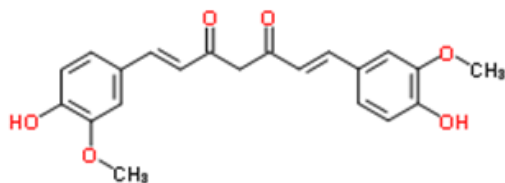


**1H22**  
**(AChE)**

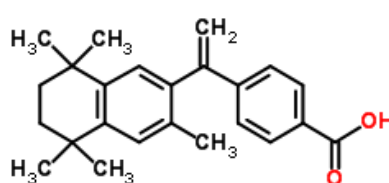


**Figure S1.** 3D structures of six targets 2MXU, 4EMA, 4K6I, 1M4H, 4Y6K, and 1H22.

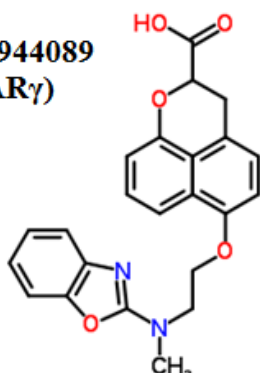
**Curcumin (A $\beta$ 11-42)**



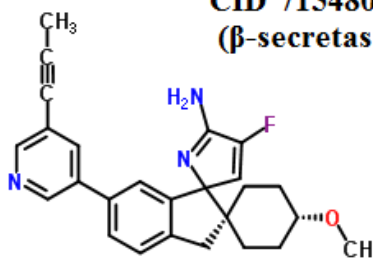
**Bexarotene (RXR $\alpha$ )**



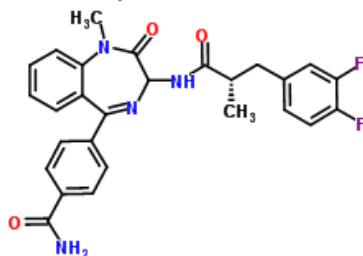
**CID 18944089  
(PPAR $\gamma$ )**



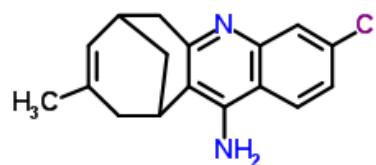
**CID 71548079  
( $\beta$ -secretase)**



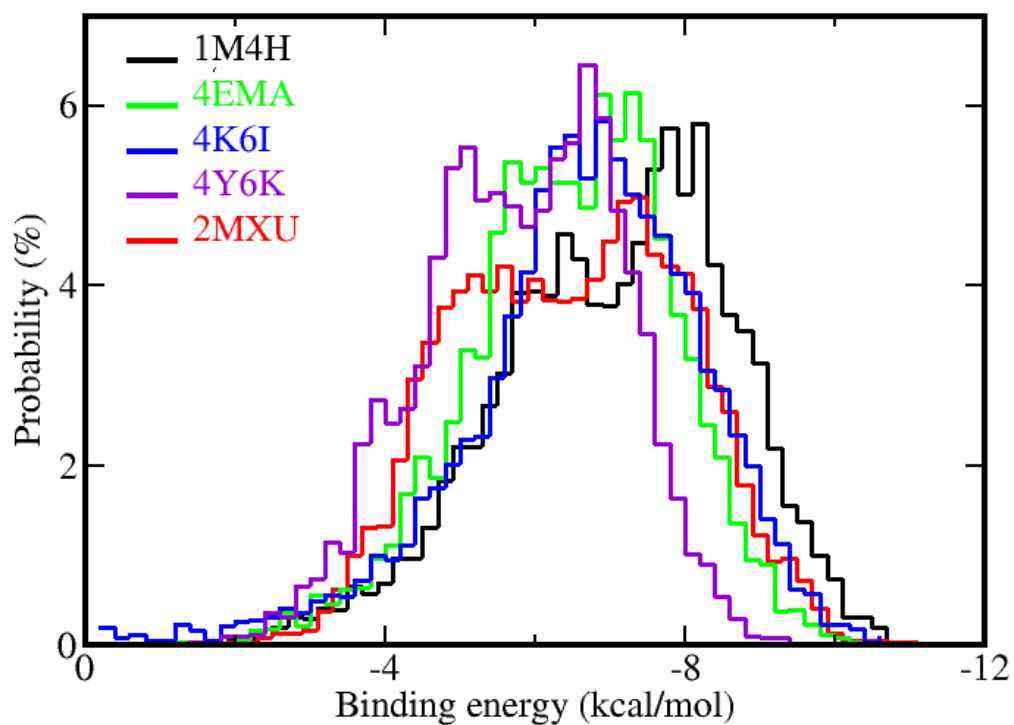
**CID 44354431  
( $\gamma$ - secretase)**



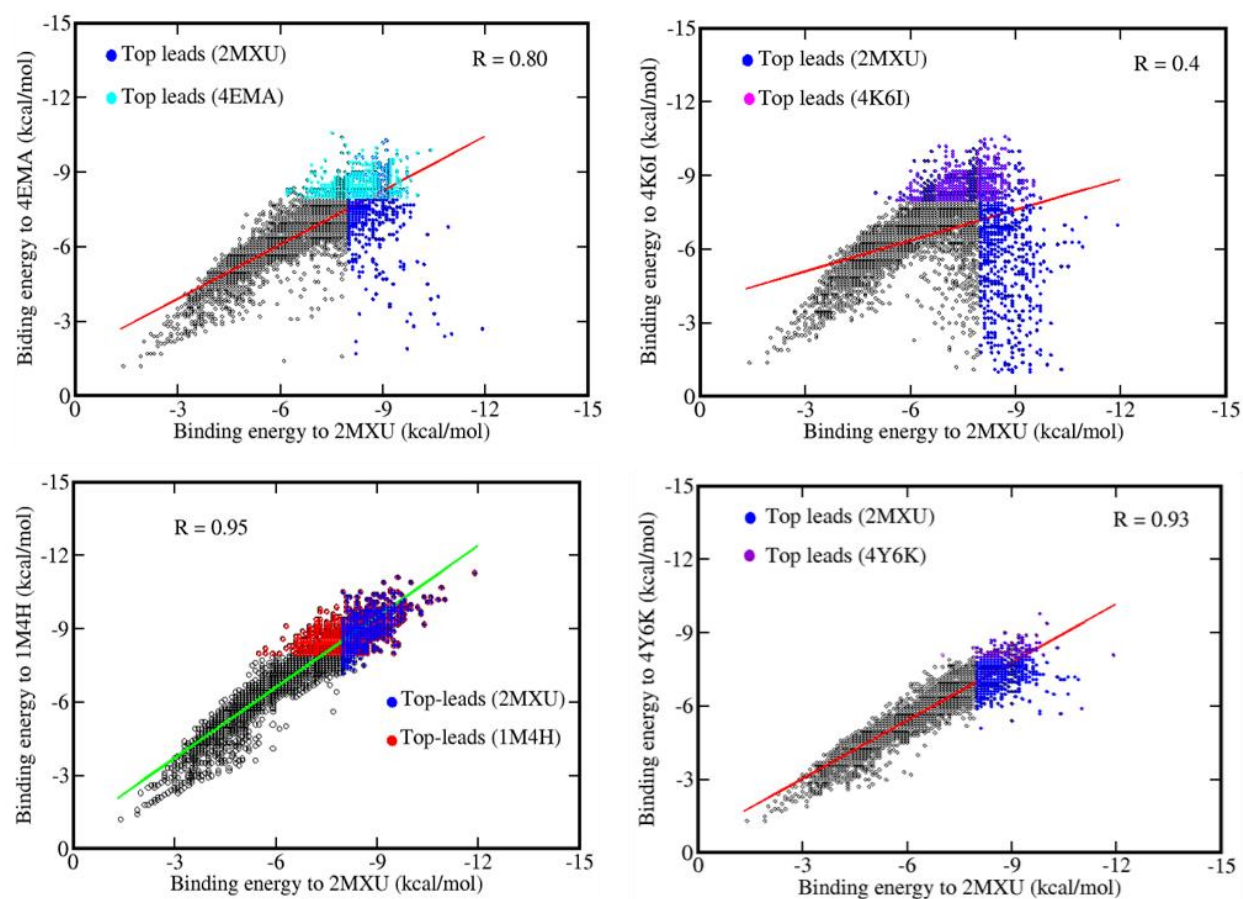
**CID 10107976  
(Acetylcholinesterase)**



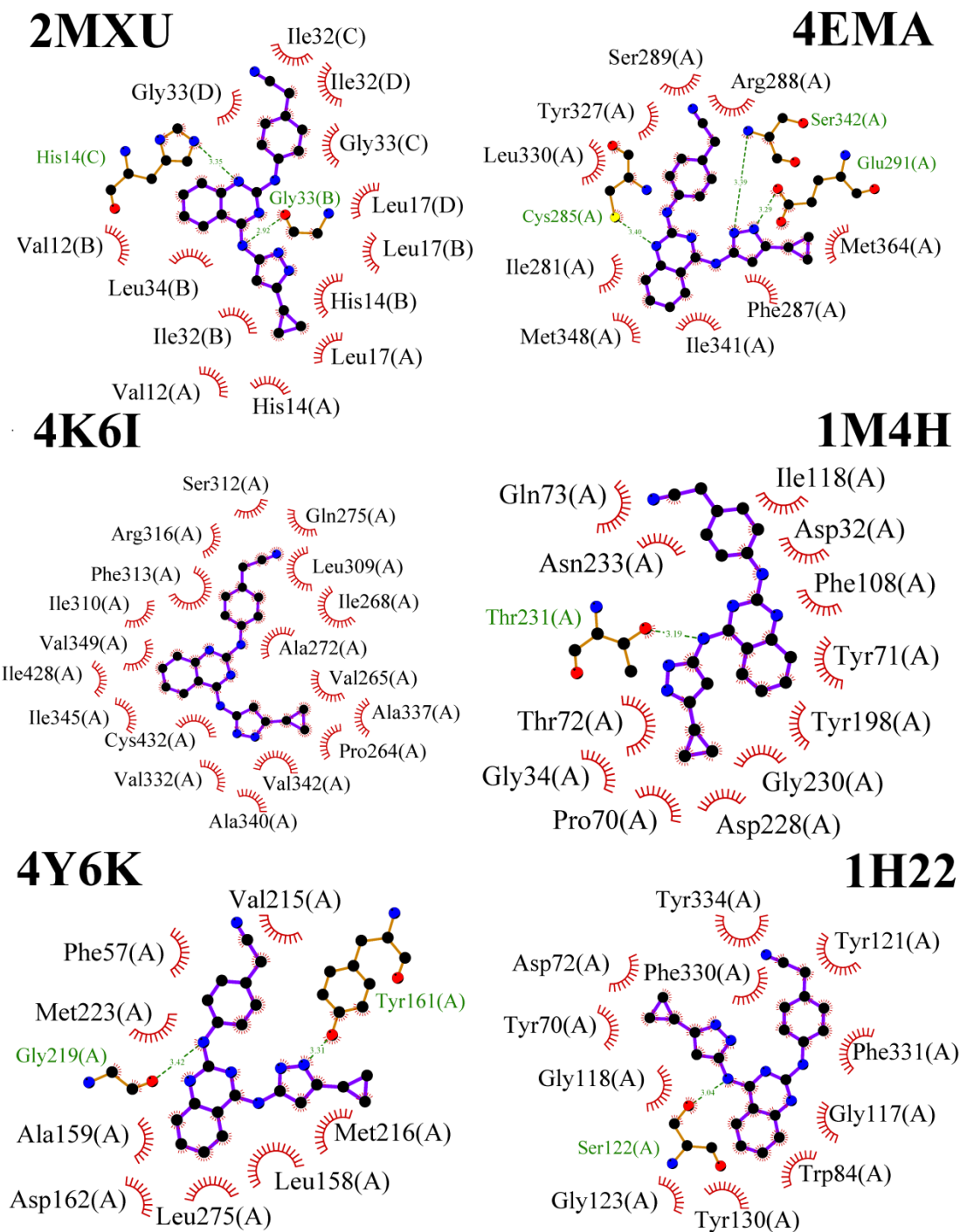
**Figure S2.** 2D structures of six reference compounds. The name of the corresponding target is also shown.



**Figure S3.** Distributions of  $\Delta E_{\text{bind}}$  of 5732 ligands to five receptors. Results have been obtained by the docking method.



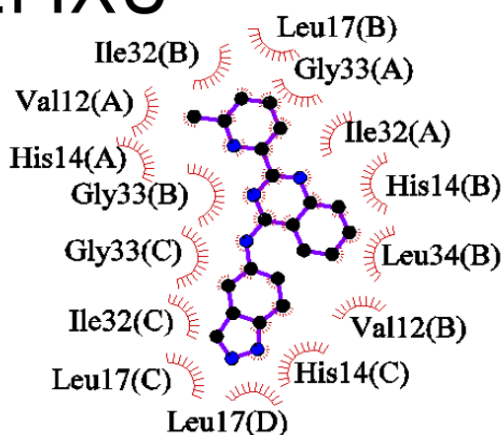
**Figure S4.** Relationship between binding energies to 2MXU and other receptors.



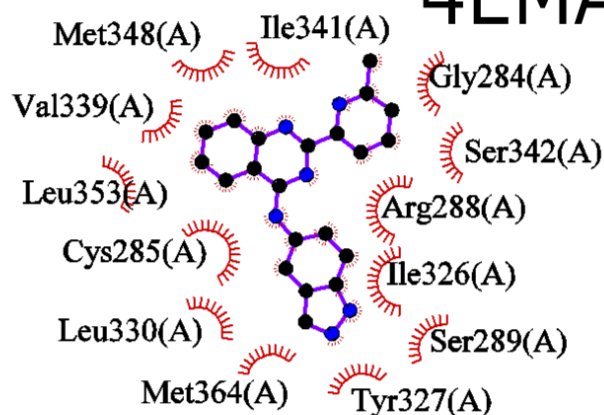
**Figure S5.** Networks of non-bonded contacts (arc) and HBs (green dashed line) of CID 16040294 with six targets. Results were provided by the docking simulation. The capital letter in the brackets refers to either monomer (A $\beta$  fibril) or chain of proteins. The plot was prepared by LigPlot+ version 1.4.4.



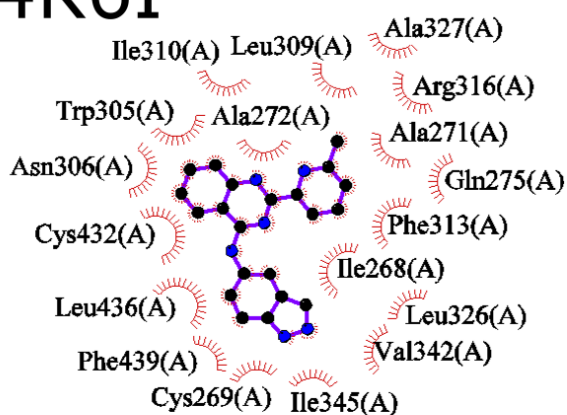
2MXU



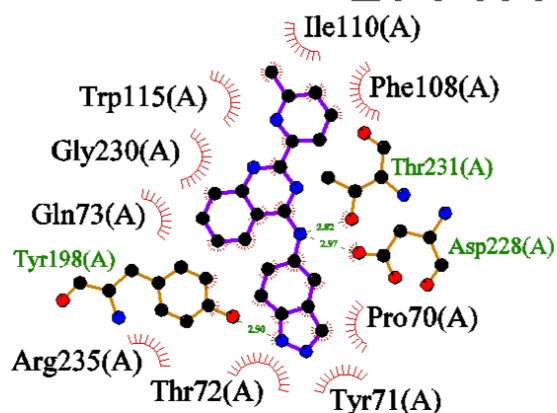
4EMA



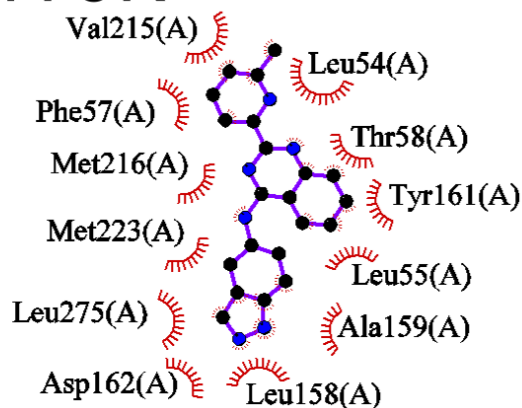
4K6I



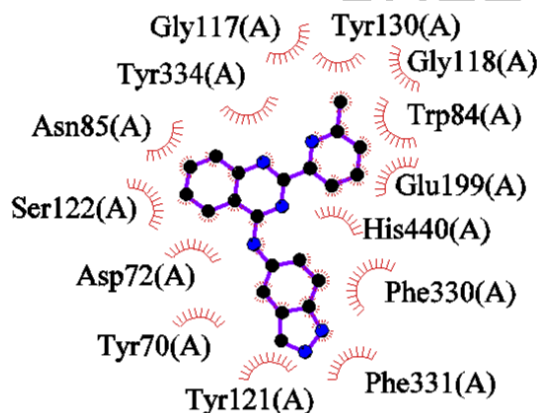
1M4H



4Y6K

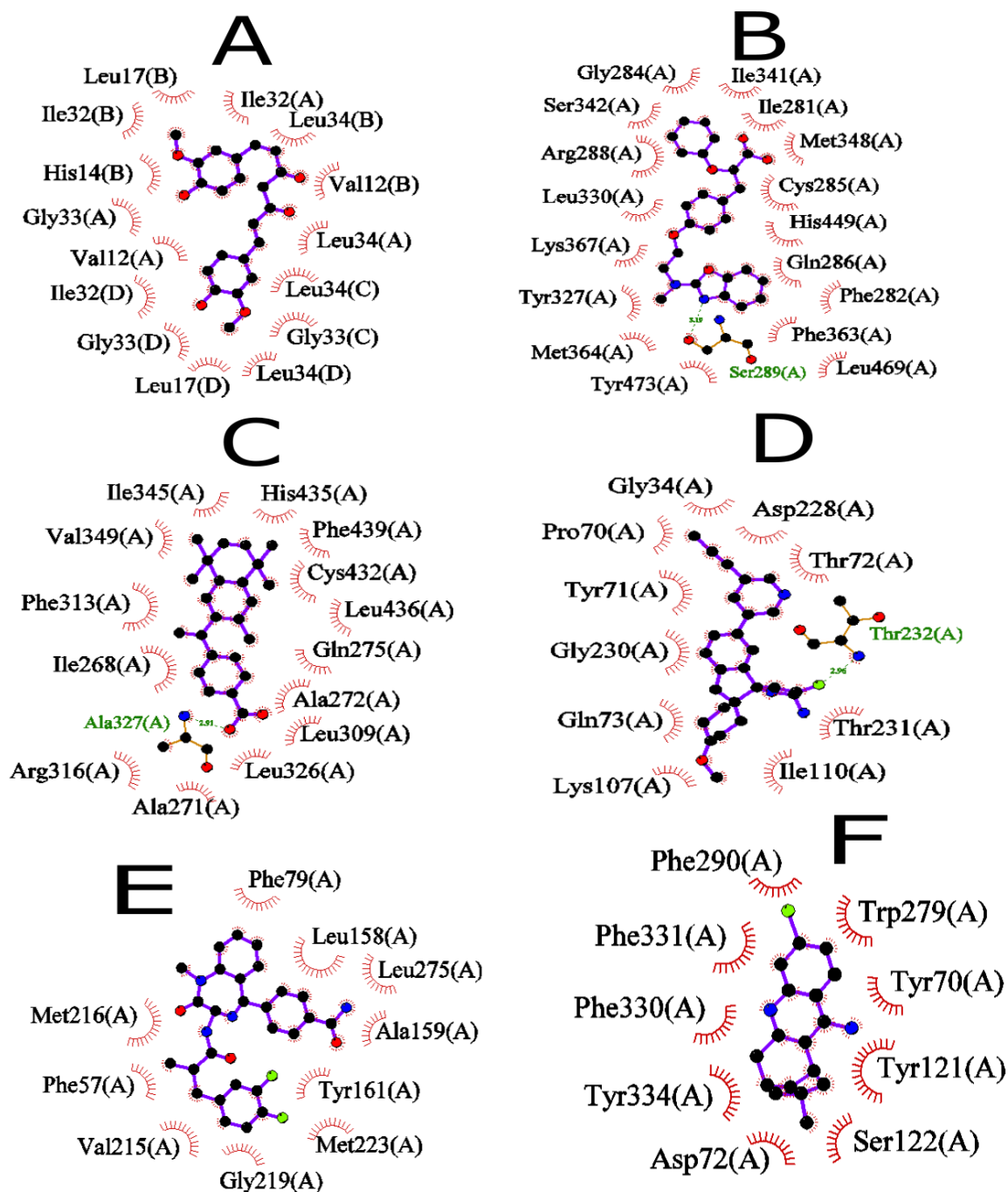


1H22



**Figure S6.** The same as Figure 3 in the main text but for compound 9998128.





**Figure S7.** Networks of non-bonded contacts (arc) and hydrogen bonds (green dashed line) in the best docking mode. The capital letter in the brackets refers to either peptide (A $\beta$  fibril) or chain of proteins. (A) 2MXU and curcumin, (B) 4EMA and CID 18944089, (C) 4K6I and bexarotene, (D) 1M4H and CID 71548079, (E) 4Y6K and CID 44354431, and (F) 1H22 and CID 10107976

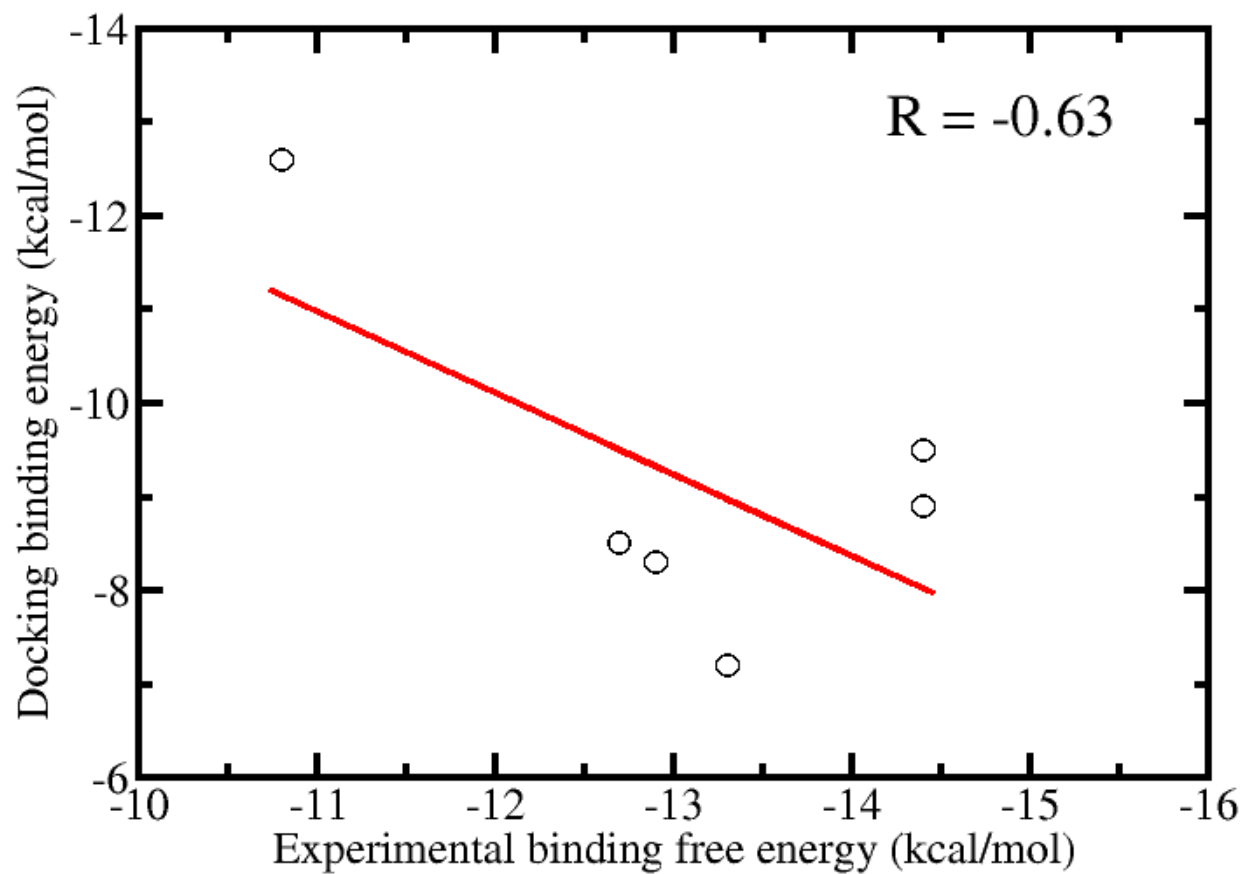
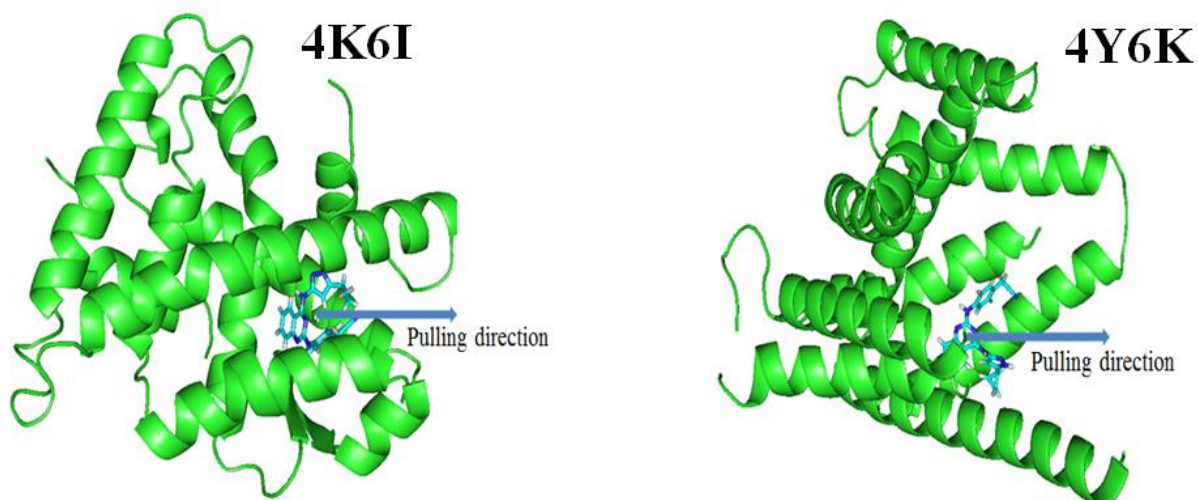
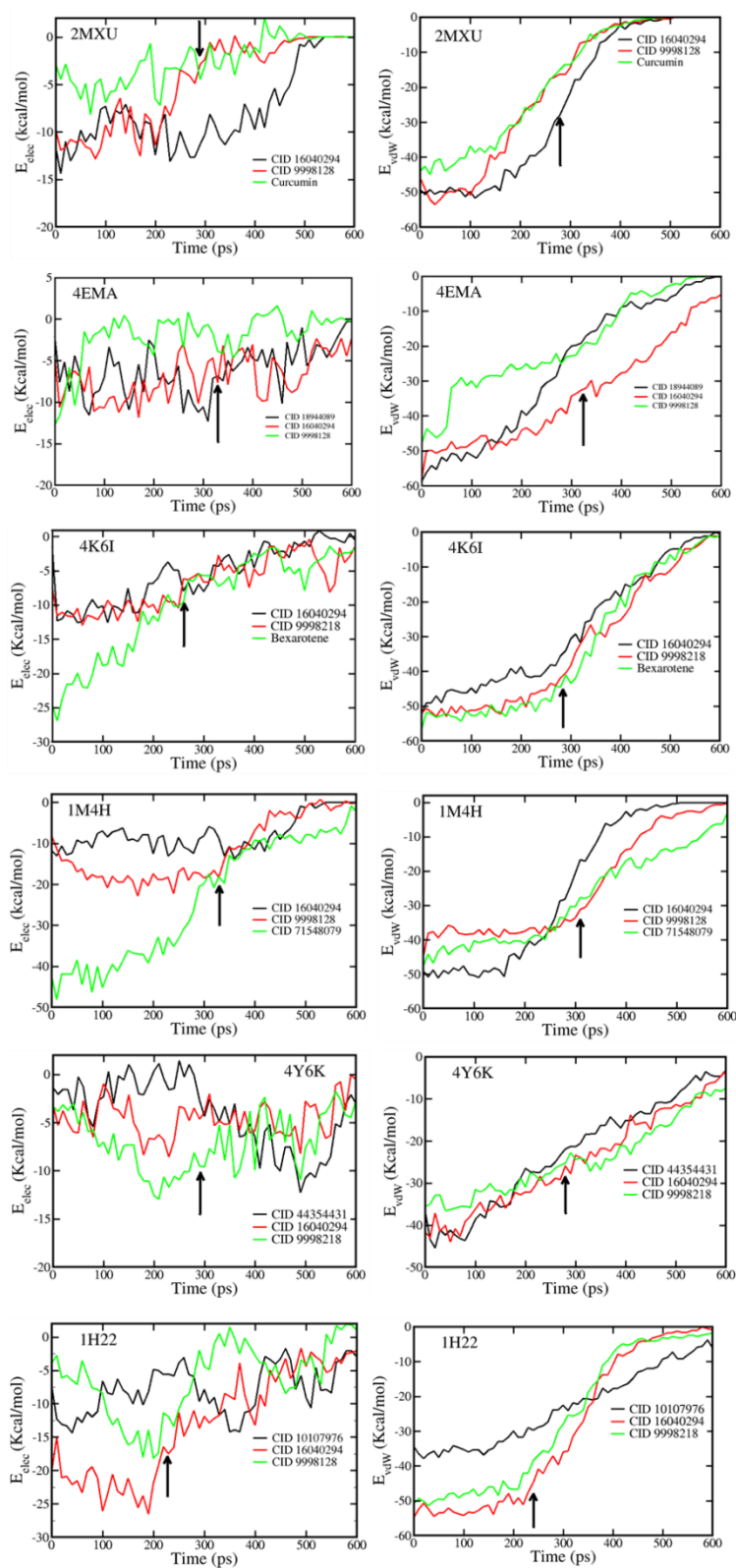


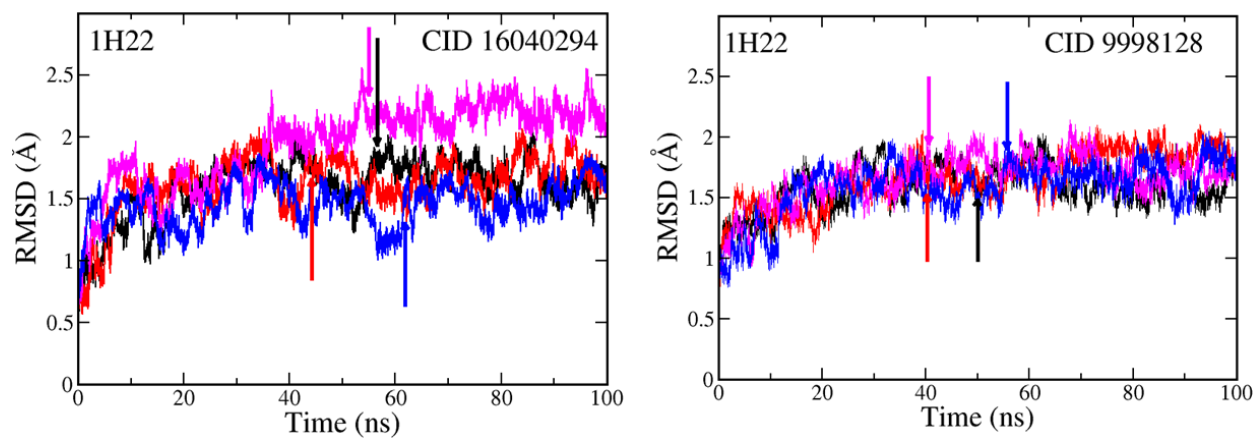
Figure S8. Correlation between docking binding energies and experimental free energies. The correlation level  $R=0.63$ .



**Figure S9.** Pulling direction of CID 998128 from the binding site of 4K6I and 4Y6K was determined by Caver 3.0, Pymol plugin.



**Figure S10.** Time evolution of vdW and electrostatic interaction energies between receptor and ligand during SMD simulation for typical complexes. Arrows roughly refer to time when the ligand quits the binding site.



**Figure S11.** Time dependences of RMSD of 1H22 in complex with CID 16040294 and CID 9998128. Arrows refer to time when the systems reach equilibrium in four independent MD runs.