

Ligand binding to anti-cancer target CD44 investigated by molecular simulations

Tin Trung Nguyen¹ · Duy Phuoc Tran² · Pham Dinh Quoc Huy^{1,3} · Zung Hoang⁴ · Paolo Carloni⁵ · Phuc Van Pham⁶ · Chuong Nguyen^{7,8} · Mai Suan Li³

Received: 8 March 2016 / Accepted: 31 May 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract CD44 is a cell-surface glycoprotein and receptor for hyaluronan, one of the major components of the tumor extracellular matrix. There is evidence that the interaction between CD44 and hyaluronan promotes breast cancer metastasis. Recently, the molecule F-19848A was shown to inhibit hyaluronan binding to receptor CD44 in a cell-based assay. In this study, we investigated the mechanism and energetics of F-19848A binding to CD44 using molecular simulation. Using the molecular mechanics/Poisson Boltzmann surface area (MM-PBSA) method, we obtained the binding free energy and inhibition constant of the complex. The van der Waals (vdW) interaction and the extended portion of F-19848A play key roles in the binding affinity. We screened natural products from a traditional Chinese medicine database to search for

CD44 inhibitors. From combining pharmaceutical requirements with docking and molecular dynamics simulations, we found ten compounds that are potentially better or equal to the F-19848A ligand at binding to CD44 receptor. Therefore, we have identified new candidates of CD44 inhibitors, based on molecular simulation, which may be effective small molecules for the therapy of breast cancer.

Keywords Breast cancer · CD44 · F-19848A · Hyaluronan · Steered molecular dynamics · Traditional Chinese medicine

Abbreviations

BCSCs Breast cancer stem cells

Electronic supplementary material The online version of this article (doi:10.1007/s00894-016-3029-6) contains supplementary material, which is available to authorized users.

✉ Paolo Carloni
p.carloni@fz-juelich.de

✉ Phuc Van Pham
pvphuc@hcmus.edu.vn

✉ Chuong Nguyen
nguyenhahungchuong@tdt.edu.vn

✉ Mai Suan Li
masli@ifpan.edu.pl

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

² University of Technology, Vietnam National University–Ho Chi Minh City, 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam

³ Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland

⁴ Center for Molecular and NanoArchitecture (MANAR), Vietnam National University–Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

⁵ Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Juelich, Juelich, Germany

⁶ Stem Cell Research and Application Laboratory, University of Science, Vietnam National University, Ho Chi Minh City, Vietnam

⁷ Theoretical Physics Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁸ Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam

HA	Hyaluronan
MM-PBSA	Molecular mechanics Poisson Boltzmann surface area
vdW	Van der Waals
TCM	Traditional Chinese medicine
MD	Molecular dynamics
SMD	Steered molecular dynamics
SC	Side chain
HB	Hydrogen bond

Introduction

Breast cancer is the most common cause of cancer-related mortality among women worldwide [1, 2]. Although there has been a significant reduction in global mortality from breast cancer during the past 20 years [3], more than 50 % of breast tumors still do not respond to current therapies [4]. Therefore, new and more effective approaches to treating breast cancer are needed. The hypothesis that breast cancer stem cells (BCSCs) are present in all malignant breast tumors is widely accepted by the scientific community [5]. BCSCs contribute significantly to chemo-resistance and relapse, and are considered to be the “cell of origin” for primary tumors, metastases, and relapses in breast cancer patients [6]. Therefore, targeting BCSCs, e.g., via small molecules, may represent a promising approach to inhibit and/or treat breast cancer.

BCSCs were first identified in human breast tumors by Al-Hajj et al. in 2003 [7]. The authors discovered that tumorigenic cells were derived from the CD44⁺CD24^{low} lineage in eight of the nine patients studied. As few as 100 of the tumorigenic cells could form tumors in NOD/SCID mice, while tens of thousands of cells expressing different cell surface markers failed to form tumors in the same mouse model [7]. The CD44⁺CD24[−] phenotype is still used today to identify and isolate BCSCs. Moreover, recent studies have suggested that CD44 down-regulation suppresses breast tumor growth in mice [8]. Therefore, inhibiting CD44 protein may represent a promising strategy for breast cancer therapy.

To develop new inhibitors of CD44, it is critical to understand the ligand binding mechanisms of the protein in molecular detail. CD44 binds to various ligands at different intracellular and extracellular positions [9]. Some experiments have shown that the interaction between CD44 and its native carbohydrate ligand, hyaluronic acid (HA; also called hyaluronan), promotes tumor growth [10]. However, our understanding of these interactions at the molecular level is limited due to a number of challenges in obtaining atomic-resolution structures of human CD44 in its bound form with ligands. These include the difficulty of purifying the samples, the flexibility of the structure, and the membrane environment, just to name a few [11]. In such cases, molecular modeling can be a useful tool to enhance insight of detailed interactions at the molecular level.

Molecular dynamics simulation approach has been used to investigate the binding between HA and the extracellular region of murine CD44, which shares 66 % sequence identity with human CD44 [12]. It was reported that the contacts of CD44 with HA are dominated by hydrogen bonds with a minor contribution from hydrophobic interactions and salt bridges. Other computational studies have suggested that the residue Tyr42 of CD44 can trigger the conformational change of CD44/HA binding domain from a low affinity to a high affinity state, including the formation of the direct contact between residue Arg41 and HA [13, 14]. Other amino acids in close proximity to the HA-binding site and water molecules have also been suggested to play some important roles in CD44 binding with HA [14–16]. Along with HA, F-19848A has also been found to bind to the extracellular regions of CD44. The F-19848A molecule can inhibit CD44 activity (IC₅₀ of 23.5 μM) [17].

Since the binding mechanism of CD44 and ligand has not been well-studied, we aimed to investigate CD44/ligand binding at the atomic level. Using the molecular mechanics/Poisson Boltzmann surface area (MM-PBSA) method, we obtained the inhibition constant of F-19848A for CD44, which is in accordance with the experiments of Hirota-Takahata et al. [17]. Our data show that binding affinity is primarily governed by van der Waals (vdW) interactions between the ligand and receptor. Residues Arg41, Tyr42, Thr76, Cys77, Arg78, Tyr79, Ile96, Cys97, Ala98, Leu107, and Tyr114 in the receptor binding cavity and the first 32 atoms of F-19848A are the most critical for binding.

Because the inhibition constant of F-19848A is relatively high (μM), there is a need to find better candidates to block CD44 activity. Currently, there are several therapeutics available in the market to treat various types of cancer, but no drug is fully effective and safe. A major problem in cancer chemotherapy is the toxicity of the established drugs. However, nutraceuticals that are natural products or extracts have been proven to be safe and effective for the treatment and management of some cancers [18]. Therefore, we screened potential lead compounds for CD44 inhibitors from natural products deposited in the traditional Chinese medicine (TCM) database [19]. Combining pharmaceutical requirements with docking and molecular dynamics simulations, we found ten compounds which are potentially more or equally effective as F-19848A at inhibiting CD44 activity. The binding mechanisms of the top ten hit compounds were evaluated at the atomic level in this study.

Materials and methods

Receptor and ligand structures

The crystal structure of CD44 was taken from Protein Data Bank (PDB ID 1POZ) [20] and was used as the receptor

throughout all molecular simulations. The 3D structure of F-19848A [17] was first built by GaussView5, and then optimized by GAUSSIAN package [21]. Structures of ligands were taken from the TCM Database [19].

Molecular docking

All ligands were docked into the binding site of CD44. PDBQT files were prepared for the receptors and the ligands using AutodockTools [22]. The Autodock Vina was employed for estimating binding energies [22]. The exhaustiveness of the global search was set to 100, and the maximum energy difference between the best and the worst binding modes was chosen as 7 kcal mol⁻¹. Ten binding modes were generated starting from random configurations of the ligand, which has fully flexible torsion degrees of freedom. The center of the grid was placed at the center of the receptor binding site [23], with grid dimensions of (23, 29, 24) Å, which is large enough to cover the receptor binding site.

Molecular dynamics simulations

The structure of receptor–ligand complex obtained in the best docking mode was used as the starting configuration for molecular dynamics (MD) simulations. The GROMACS package [24] was used to carry out MD simulations. The topology and coordinate files of F-19848A and other ligands were generated by PRODRG 2.5 Web Server [25]. Parameters of CD44 (atom types, charges, etc.) were assigned by GROMOS43a1 force field [26], while parameters of ligands were assigned by PRODRG 2.5 server. The systems consisted of 31435 to 39020 atoms. Three-site water model SPC [27] was used to fill the simulation box. The water box was set in such a way that the complex was placed at least 1 nm from the box edge. For the CD44 and F-19848A complex, for instance, the box size was set to 7.29 × 7.29 × 7.29 nm³, and nine Na⁺ ions were added to neutralize the system.

The leapfrog algorithm [28] was used to integrate the equations of motion with a time step of 2 fs. The LINCS algorithm [29] was employed to constrain the length of all bonds. Electrostatic and vdW interactions were calculated with a cut-off of 1.0 nm. Particle-mesh Ewald method [30] was used to treat the long-range electrostatic interactions. First, the steepest decent method was used for minimizing the energy of the system. Then a 500 ps NVT run was performed at a temperature of 300 Kelvin (K), which was kept constant by the modified Berendsen (V-rescale) temperature coupling method [31]. Four hundred ps of NPT ensemble with Berendsen pressure coupling [32] and 4 ns of NPT ensemble with Parrinello-Rahman barostat [33] at 1 atm were performed successively to stabilize the system. The production NPT MD simulations were then carried out at the same temperature and pressure for 50 ns.

Steered molecular dynamics (SMD)

Recently, the steered molecular dynamics (SMD) method [34] has been discovered to be a useful tool for drug design [35]. The simulation box sizes for SMD simulation were varied depending on the pulling direction of ligands from the receptor. To set the box, we first determined possible pathways of ligands using the software CAVER 3.0 [36] and Pymol plugin, and chose the easiest path for the ligand to get away from the receptor as the pulling direction. The receptor was rotated in such a manner that the binding site was toward the pulling direction along the z-direction (Fig. 1).

The position-restrained MD simulation was performed for 100 ps to allow water molecules to move into the binding site. The system was gradually heated up to 300 K during 500 ps of MD simulation followed by 500 ps in NVT ensemble with modified Berendsen (V-rescale) temperature coupling. Then 2 ns simulation in NPT ensemble with Parrinello-Rahman pressure coupling was performed to ensure that the system was stable. In the last step, we performed the SMD simulations. The force was applied to the center of mass of the ligand. The spring constant of cantilever was chosen as $k = 600 \text{ kJ (mol nm}^2)^{-1}$, which is a typical value used in atomic-force microscope (AFM) experiments [37]. The loading speed was set as $v = 0.005 \text{ nm/ps}$ [38]. Five independent trajectories were run with different random seeds. To completely pull the ligand out of the receptor, we performed an SMD run for 400 ps after 100 ps of NPT ensemble equilibration.

In this study, rupture forces and pulling works were calculated to rank the binding affinity of the compounds with CD44. Rupture force is the maximum value of the force during the pulling process. It is considered as the value when a ligand starts to unbind and leave the binding site. Pulling work under the action of a force, $F(t)$, is calculated by the following formula:

$$W_{pull} = \int_0^{x_{max}} \vec{F} \cdot d\vec{x} \approx \int_0^{t_{pull}} F(t) \cdot v \cdot dt \quad (1)$$

where v is the loading speed, dt is the timestep, $F(t)$ is value of force at time t , taken from the force-time profile of each trajectory.

Molecular mechanics Poisson-Boltzmann surface area method

We applied the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method [39] to estimate the binding affinity of F-19848A and top hit compounds screened from the TCM Database. Overall, in the MM-PBSA approach, the binding free energy of ligand to the receptor is defined as follows:

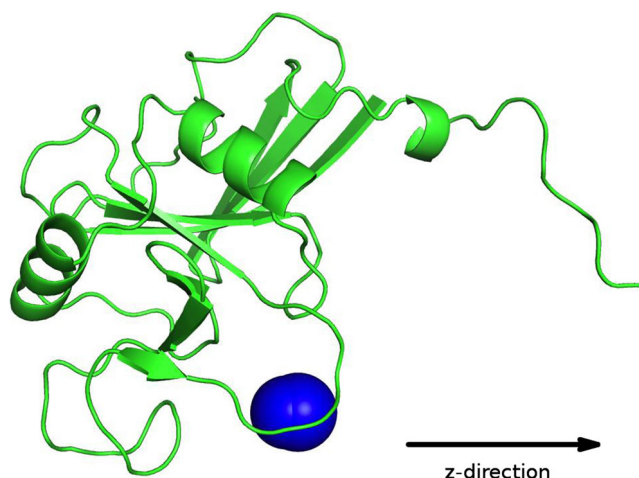


Fig. 1 The pulling direction of F-19848A from CD44. Using CAVER 3.0 we found three pathways for F-19848A. The z-direction is the easiest pathway (lowest rupture force) for ligand to escape from the binding pocket

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

where ΔE_{elec} and ΔE_{vdw} are contributions from electrostatic and vdW interactions; ΔG_{sur} and ΔG_{PB} are nonpolar and polar solvation energies. The entropic contribution $T\Delta S$ is estimated using the quasi-harmonic approximation [40]. Snapshots collected in equilibrium were used to compute the binding free energy. The APBS software package [41] was used for calculating the polar solvation energies by solving the corresponding linear Poisson-Boltzmann equation.

Results and discussion

MD simulation for CD44 and F-19848A complex

Binding affinity of F-19848A To study the binding mechanism of F-19848A to receptor CD44, we performed four independent 50 ns MD simulations. As evident from the time dependence of root mean square deviation (RMSD) of the receptor, the complex reached equilibrium at different times, depending on MD trajectories (Fig. S1 in Supporting information (SI)). Snapshots collected in the last 10 ns of four MD trajectories were used to calculate the number of the side-chain (SC) contacts and hydrogen bonds (HBs). The SC contact map shown in Fig. 2a provides detailed information about the CD44-F-19848A interactions. As depicted, 11 residues of CD44, including Arg41, Tyr42, Thr76, Cys77, Arg78, Tyr79, Ile96, Cys97, Ala98, Leu107, and Tyr114, are among the most important residues as they have a mean number of contacts exceeding five (Fig. 2b). Interestingly, the experiment [23] revealed that eight of them (Arg41, Tyr42, Cys77, Arg78, Tyr79, Ile96, Cys97, and Ala98) make prominent contact with HA. From the plot for per-atom contacts (Fig. 2c), we

observed that the extended part of F-19848A, consisting of atoms 1–32, contribute dominantly (about 60 %) to the SC interaction. This suggests that F-19848A may compete with HA upon binding, via its extended part, since F-19848A binds to CD44 in the same region. The atom indices of F-19848A are listed in Fig. 3. Atoms 60–68 appear to make an important contribution to the SC interaction with the receptor (Fig. 2c).

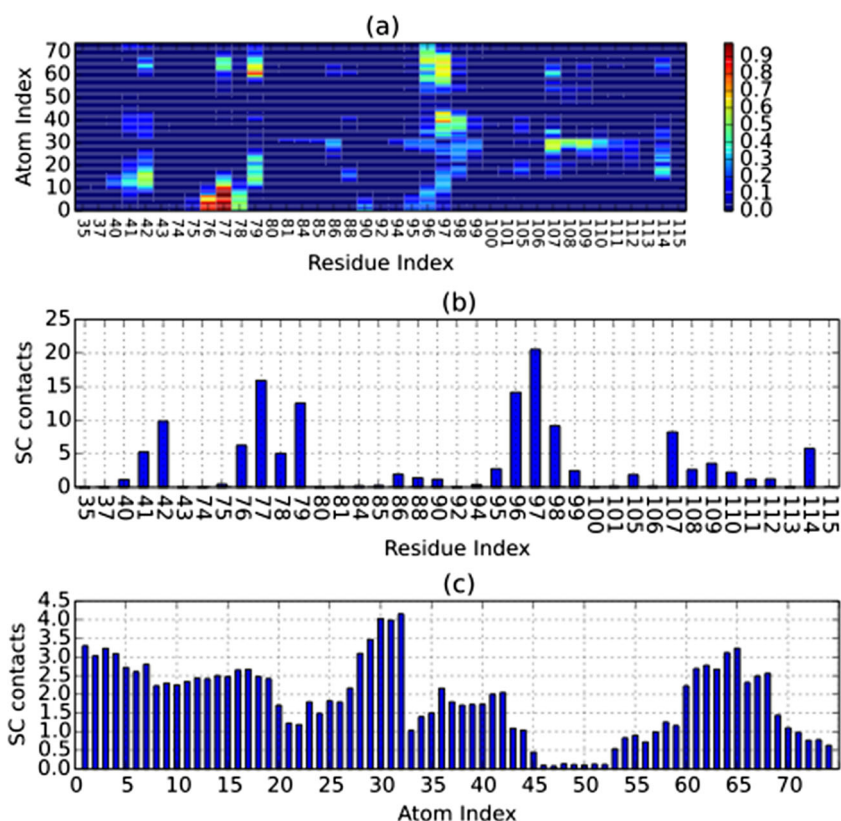
The HB contact map (Fig. S2 in SI) is much poorer than the SC one. Only two residues (Thr76 and Cys77) contribute substantially to the HB network (Fig. S2b and S2d in SI), while the $-\text{COO}$ group of F-19848A (Fig. 3) shows a high propensity for hydrogen binding with CD44 (Fig. S2c in SI). Therefore, HBs presumably play a minor role in the interaction between CD44 and F-19848A. Note that during the equilibrium MD simulation, the ligand experiences conformational changes in its binding position leading to the typical hydrogen bond network (Fig. S2d), which is different from the docking one (Fig. 9).

Using Eq. 1 and snapshots collected in the last 10 ns, we obtained the binding free energy (Table S1 in SI). The mean binding energy of F-19848A obtained by the MM-PBSA method is $-12.1 \pm 1.6 \text{ kcal mol}^{-1}$. The experimentally measured inhibition constant IC50 is about 23.5 μM [13]. On the other hand, ΔG_{bind} may be estimated from IC50 using the formula $\Delta G_{\text{bind}} \approx RT \ln(\text{IC}_{50})$, where $R = 1.987 \times 10^{-3} \text{ kcal mol}^{-1}$, $T = 300 \text{ K}$, and IC50 is measured in moles. Thus, $\text{IC}_{50} = 23.5 \mu\text{M}$ roughly corresponds to $\Delta G_{\text{bind}} = -9 \text{ kcal mol}^{-1}$. The error bars reflect whether the theoretical predictions is far or not from the experimental results. As expected for a ligand with a long hydrophobic tail (Fig. 3), the vdW contribution to ΔG_{bind} is large (Table S1). Therefore, the vdW interaction plays a key role in the binding affinity of F-19848A to CD44. Using the ff03 force field [42], Samsonov et al. have reported that the vdW interaction is dominant in the binding between CD44 and HA [43]. Thus, the vdW interaction is presumably a major driving force in the competition of F-19848A or HA binding to CD44.

The vdW interaction Since the F-19848A-CD44 complex is maintained predominantly by vdW interaction (Table S1), we investigated the vdW interaction in greater detail. Residues Tyr42, Cys77, Tyr79, Ile96, Cys97, Leu107, and Tyr114 contribute to the vdW energy remarkably better than the other residues (Fig. 4). As expected, these top residues also form many SC contacts with the ligand (Fig. 2). Thr76, Cys77, Ser95, and Ile96 contribute substantially not only to the vdW interaction but also to hydrogen bonding (Fig. 4 and Fig. S2 in SI).

The contributions of individual ligand atoms of F-19848A to the vdW interaction energies are depicted in Fig. 5. Consistent with the analysis on the SC map, the part of F-19848A which expands over atoms 1–32 interacts strongly with the receptor. The contribution from this extended part,

Fig. 2 (a) Map of SC contacts between CD44 and F-19848A. Bar color refers to population of contact during MD runs in equilibrium. (b) Per-residue SC contacts. Eleven residues of CD44 (Arg41, Tyr42, Thr76, Cys77, Arg78, Tyr79, Ile96, Cys97, Ala98, Leu107, and Tyr114) have populations exceeding 5 %. (c) Per-atom SC contacts. Thirty-two atoms of F-19848A have indices from 1 to 32 which contribute dominantly (approximately 60 %) to the contact network



which comprises about 65 % of the structure, suggests it has a central role in the binding affinity of F-19848A to CD44. Atoms O41, C60, O61, C63, O64, C66, and O67 from the left and right aromatic rings also strongly interact with ligand via vdW interaction (Figs. 3 and 5), but the contribution of the middle ring is less important.

Identification of novel CD44 ligands

We considered 32364 compounds derived from Eastern herbs and plants which were deposited in the TCM database [19] (website: <http://tcm.cmu.edu.tw>). Applying Lipinski's rule of five [44] to this set we obtained 3699 ligands that have drug-

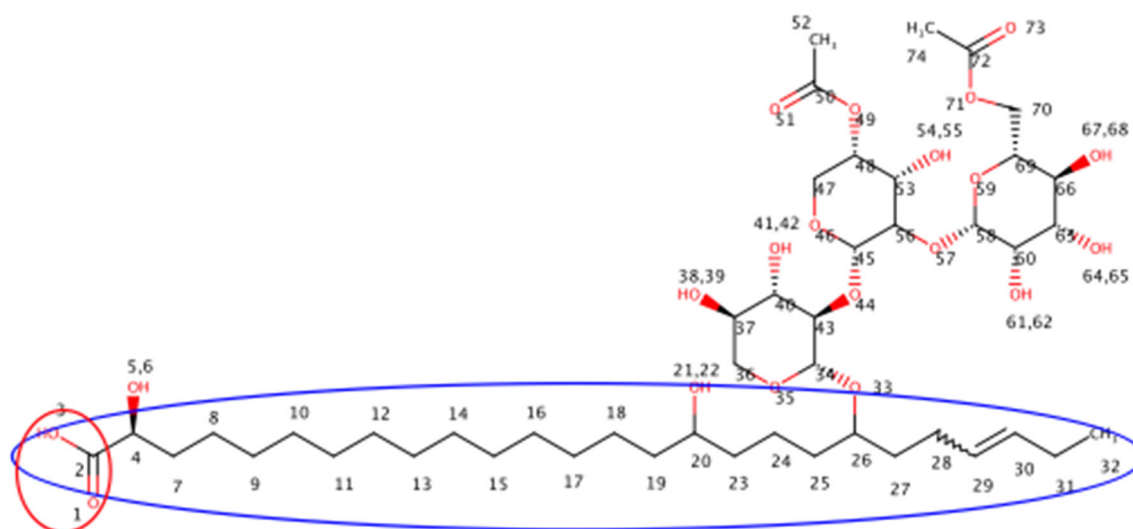
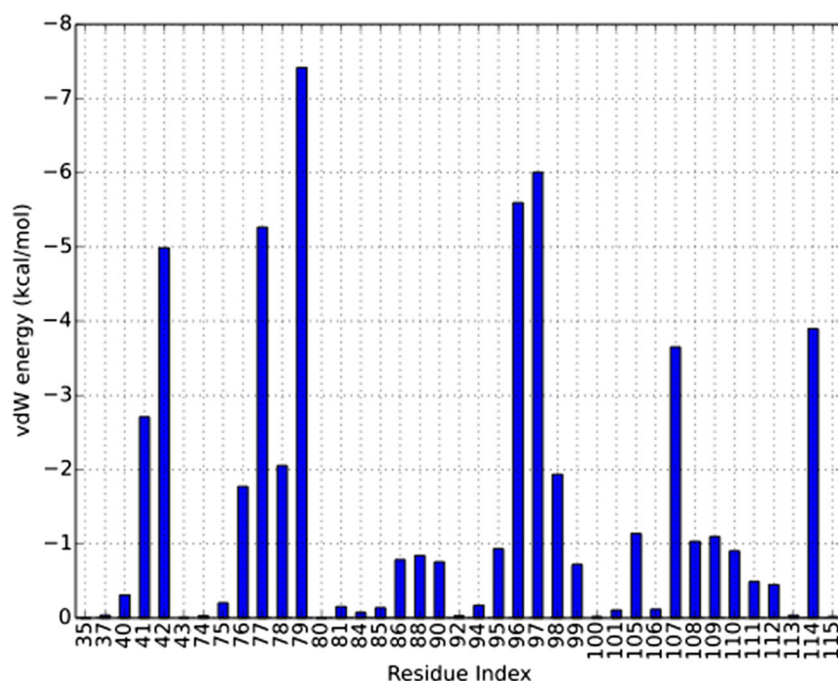


Fig. 3 2D structure of F-19848A. The extended portion of F-19848A which interacts strongly with its receptor is marked by the blue circle. The red circle represents -COOH which becomes -COO in water. The -COO group plays a central role in forming hydrogen bonds with the receptor

Fig. 4 Per-residue distribution of the vdW interaction energy. Results were obtained in the last 10 ns of four MD runs



like properties. It is important to consider that many natural products remain bioavailable, despite violating the rule of five [45]. Thus, not all pharmaceutically relevant compounds might have been selected here.

Using Autodock VINA, we docked compounds from the reduced set to the binding site of CD44. The Top 99 compounds which had the lowest binding energy, as estimated by Autodock VINA, were selected for further SMD simulation. Figure 6 shows the dependence of force experience by ligand on time and position for the three ligands: F-19848A,

25501, and 26440. Compound 26440 and 25501 were selected for comparing with F-19848A; the former is mechanically less stable, with a lower rupture force, while the latter is more stable (Fig. 6). For compound 25501, the force reached a maximum $F_{\max} \approx 700$ pN approximately at 220 ps. In the case of F-19848A and compound 26440, the rupture happened earlier at about 160 and 140 ps with lower maximum forces of 550 and 470 pN, respectively. The rupture positions of 25501, F-19848A, and 26440 were, respectively, 0.4 nm, 0.24 nm, and 0.21 nm away from the initial position,

Fig. 5 Contributions of individual atoms of F-19848A to the vdW interaction energy with CD44. Results were obtained in the last 10 ns of four MD runs

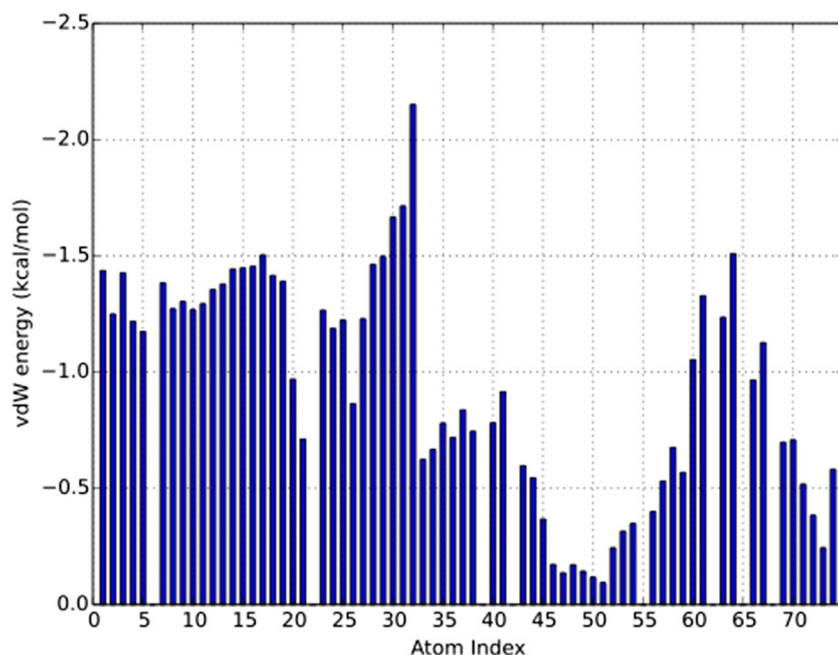
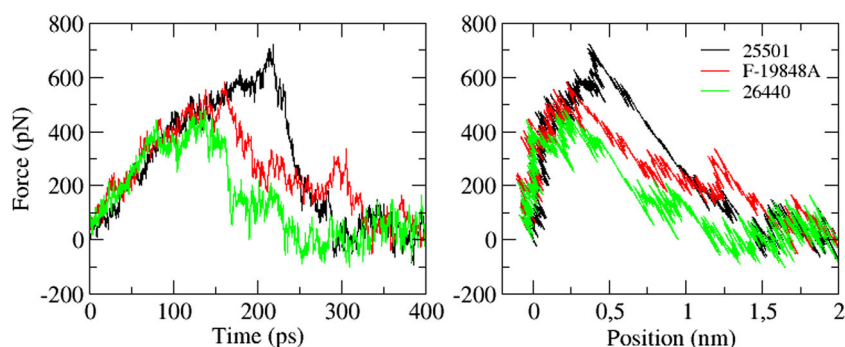


Fig. 6 Force-time and force-position profiles for 25501, F-19848A, and 26440



suggesting that there are slight conformation changes before the dissociation occurs.

The evolution of the steering force with displacement was initially linear, then nonlinear before the rupture (Fig. 6). At that stage, the ligand stayed in the binding pocket. After achieving the maximum value in the profiles, the force decreased steeply and fluctuated around zero after the ligand detached from the receptor along the pulling direction.

Ranking of ligands by rupture force F_{\max} and pulling work W_{pull} We performed five independent SMD runs for each of the 99 ligands. The force-time profiles for five trajectories for F-19848A (Fig. S3 in SI) showed that F_{\max} is not hypersensitive to MD runs. This observation remains valid for other compounds. The mean values of rupture forces and pulling works were calculated (Fig. 7). The correlation between these two values was very high ($R = 0.86$) (Fig. S4 in SI). Both values were therefore used to rank the binding affinity of ligands with CD44 as they were expected to be more reliable than the docking method [46]. For instance, compound 25501 was only 53rd in docking stage, but it became the most potent in rupture force and pulling work ranking.

Our data show that, among 99 compounds, there are ten compounds that have rupture forces higher or as good as the compound F-19848A (Table 1). Interestingly, these ten compounds possess at least two rings which substantially stabilize the complex with CD44. From analysis with the same SMD data, we have found that there are only three compounds that consume the pulling work greater or equal to F-19848A (Fig. 7, Table S2). Note that the top three compounds revealed by pulling work were all in the list of top ten compounds revealed by rupture forces (Table S2). This suggests that although SMD is a suitable tool for hit identification, it may not always be perfect for lead optimization because of dependence on the dissociation pathway. In this event more precise methods, such as MM/PBSA, should be used to further verify the SMD results. Thus, we applied the MM/PBSA method to the top ten hits revealed by SMD.

Nature of binding of top hits revealed by SMD method To understand why, for instance, compound 25501 is better than

the others, we monitored the electrostatic and vdW interactions as a function of time (Fig. 8). The initial value of the vdW energy was about $-48 \text{ kcal mol}^{-1}$ for 25501 and F-19848A, and $-30 \text{ kcal mol}^{-1}$ for 26440, a compound that was not in the top ten list for both rupture force and pulling work ranking. Since at the beginning the electrostatic interaction energy of ligand 25501 was $-48 \text{ kcal mol}^{-1}$ versus $\approx 0 \text{ kcal mol}^{-1}$ for F-19848A, the former should bind much stronger than the latter. The total interaction energy (vdW and electrostatic) of ligand 26440 remained much higher than that of 25501 and F-19848A, resulting in the lower binding affinity of 26440. As time increased, the vdW energy almost gradually vanished due to the complete dissociation of ligand from the binding pocket, while the time dependence of the electrostatic term was much more complicated. In the case of 25501, E_{elec} fluctuated around $-48 \text{ kcal mol}^{-1}$ during the first 230 ps and then abruptly leveled up to 67 kcal mol^{-1} in the two-state manner. In the case of F-19848A, E_{elec} steadily rose from ≈ 0 to $\approx 95 \text{ kcal mol}^{-1}$. Thus, one can conclude that the repulsive electrostatic interactions assist the dissociation of these ligands. To explain this phenomenon, we considered the charge distributions of receptor and ligands. Figure S5 in SI shows the charge surface of CD44, where the binding site is a narrow region with positive charge bound by residues Arg41, His92, Ile96, and Ala99 [23], and surrounded by a negatively charged region. Since the net charge of CD44 is $-8e$, the negative charges dominate over the positive ones. Ligands 25501 and F-19846 both have a net charge of $-1e$ because the former contains a $-\text{SO}_3$ group, while the latter has a $-\text{COO}$ group (Fig. S6 in SI). In the binding site, the attraction between positive charges of CD44 and negative charges of 25501 is dominant but for F-19848A, it is nearly compensated by the repulsion between negative charges of ligand and receptor. As the ligands are pulled out of the binding site, the attractive interaction rapidly becomes weak, and the repulsion exists because the negative charges distribute dominantly. Therefore, the electrostatic energy increases as a function of time in both cases (Fig. 8).

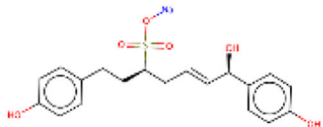
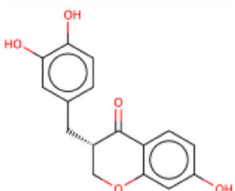
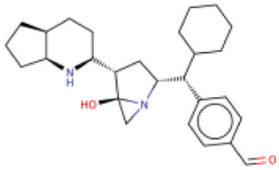
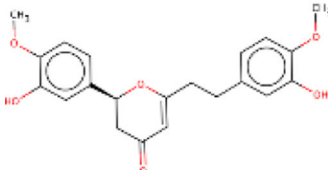
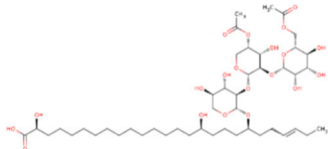
In the case of a neutral compound 26440, the electrostatic energy increases from $-14 \text{ kcal mol}^{-1}$ to zero and the vdW interaction remains dominating (Fig. 8). The stability of 25501

Table 1 The top ten ligands ranked by SMD rupture forces. Results were obtained by five independent SMD simulations. Error bars are from an average of over five runs. Ten ligands have F_{\max} higher than F-19848A.

The first column shows numbers in parentheses which refer to ranking by the docking method. ID in TCM database, IUPAC names, common names and 2D structures are shown (some of common names are not available)

SMD Rank	ID	IUPAC name	Common Name	F_{\max} (pN)	2D Structure
1 (53)	25501	Sodium (1R,5R)-1-Hydroxy-1-(4-hydroxyphenyl)-7-[4-(2-hydroxyethoxy)phenyl]hepta-2-en-5-sulfonate		713 ± 13	
2 (56)	25497	Sodium (1R,5S,7S)-1-Hydroxy-1,7-bis(4-hydroxyphenyl)nona-2-en-5-sulfonate		584 ± 25	
3 (17)	26269	N-[(E)-2-[(2S,3R)-2-acetamido-3-(3,4-dihydroxyphenyl)-2,3-dihydro-1,4-benzodioxin-6-yl]ethenyl]acetamide		570 ± 32	
4 (64)	29162	(6Z)-6-(3-oxo-1H-indol-2-ylidene)indolo[2,1-b]quinazolin-12-one	Qingdain one	568 ± 34	
5 (83)	29340	7,10,11-Trihydroxy[2]benzofuro[6,5,4-kl]xanthene-1,3-dione	Rufescidine	555 ± 38	
6 (78)	17528	(2S,3R)-2-[(E)-cinnamyl]-3,5-dihydroxy-8-(hydroxymethyl)-2-methyl-3,4-dihydropyrano[3,2-g]chromen-6-		543 ± 20	

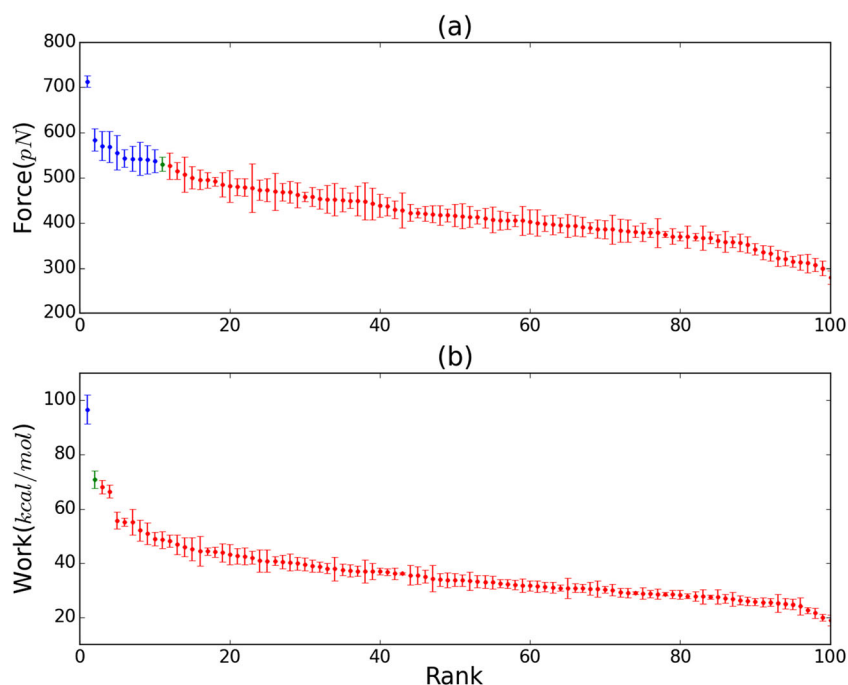
Table 1 (continued)

7 (59)	25752	sodium;(E)-7-hydroxy-1,7-bis(4-hydroxyphenyl)hept-5-ene-3-sulfonate		542 ± 28	
8 (81)	31820	(3S)-3-[(3,4-dihydroxyphenyl)methyl]-7-hydroxy-2,3-dihydrochromen-4-one	3-Deoxysap panone B	542 ± 37	
9 (63)	9973	4-[(R)-[(2R,4S,5R)-4-[(2S,4aS,7aS)-2,3,4,4a,5,6,7,7a-octahydro-1H-cyclopenta[b]pyridin-2-yl]-5-hydro		540 ± 31	
10 (31)	31515	(2S)-2-(3-hydroxy-4-methoxy-phenyl)-6-[2-(3-hydroxy-4-methoxy-phenyl)ethyl]-2,3-dihydropyran-4-one	Cyclocurcumin	537 ± 26	
11	F-19848 A			530 ± 15	

is maintained by both vdW and electrostatic interactions since their contributions are nearly equal. On the other hand, consistent with the MM-PBSA results, the contribution of the vdW interaction is dominant in the case of F-19848A. Thus, our analysis revealed that strong electrostatic interactions are what makes the champion ligand 25501 different from the other compounds.

Figure 9 shows the HB networks of 25501, F-19848A, and 26440 in complex with CD44 in the conformations which have been used as starting conformations for SMD simulations. They were obtained from the best docking mode conformations by the minimization step. Residues Tyr105, Tyr79, Ser112, Thr108, and Ser109 form seven HBs with 25501. Six out of seven HB acceptors are oxygen atoms of the -SO_3

Fig. 7 Ranking of 99 ligands by rupture force (a) and pulling work (b). The green dot refers to the reference ligand F-19848A, while blue dots refer to compounds with higher values of rupture force or work. Red dots denote compounds with values lower than F-19848A. Results were obtained from five independent SMD runs



group suggesting that this group plays an important role in maintaining the complex. Together with the electrostatic and vdW interactions discussed above, the stability of 25501 in complex with CD44 is contributed appreciably by all three factors. In the case of F-19848A, the negative carboxylate group $-\text{COO}^-$ is the most important part forming all three HBs with Ser112 and Tyr79. Compound 26440 has only two HBs with residues Ser42 and Arg78. The poor HB networks of F-19848A and 26440 complexes suggest that the hydrogen bonding does not contribute significantly to the stability of their complexes with CD44. Overall, all three factors (electrostatic interaction, vdW interaction and hydrogen bonding) play key roles in the binding affinity of 25501, while the stabilization of F-19848A and 26440 in the binding pocket of CD44 is mainly governed by the vdW interaction. This supports 25501 as a champion lead for blocking CD44 activity.

Determining binding free energies for top hit compounds using MM-PBSA method The SMD method is capable of predicting relative binding affinity but it cannot be used to compute the absolute binding free energy. On the other hand, for some cases, this method has provided good correlation with experiments [35], though the predictions are not always perfect. Therefore, we computed the absolute binding free energy, ΔG_{bind} , for the top ten compounds revealed by SMD using the MM-PBSA method [47, 48]. Parameters for MD simulations in this section were all the same as those used for the CD44 and F-19848A complex. Because ΔG_{bind} obtained by MM-PBSA for F-19848A was not very sensitive to MD runs (see SI), one MD trajectory was carried out for each top hit ligand. As shown in Fig. S7 in the SI, the studied complexes reached equilibrium at different time scales ranging from 20 to 40 ns. The binding free energies of all systems were calculated using snapshots

Fig. 8 Time dependence of electrostatic and vdW interaction energies for 25501, F-19848A, and 26440

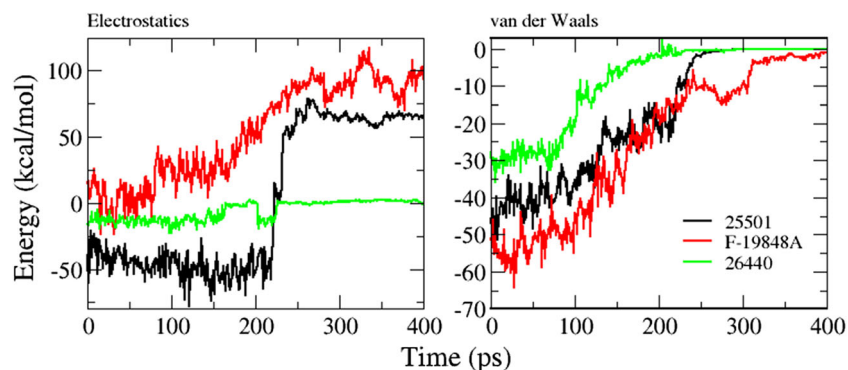
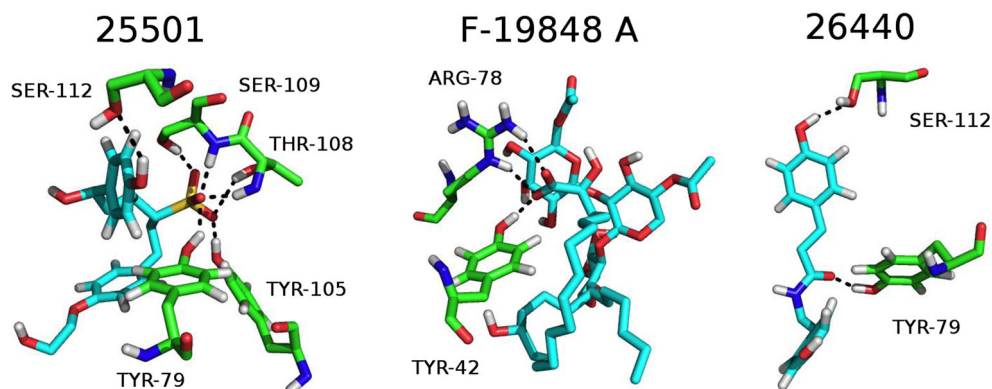


Fig. 9 HB networks of 25501, F-19848A, and 26440 systems. These conformations were used as initial conformations for the SMD simulations and are only slightly different from conformations obtained by the docking method



collected from the last 10 ns of MD trajectories. The final results are shown in Table 2. Of note, nine of the ten ligands have binding affinity higher than F-19848A, which has almost the same binding strength as compound 31820. Thus, both MM-PBSA and SMD method support these ten compounds as good candidates for blocking activity of CD44. Complementary to previous works [35, 38, 49–51], our results imply that the SMD method is useful in drug design.

Two compounds, 9973 and 29162, were selected for further analysis because they bound tightly to CD44 but with different mechanisms. The binding of 9973 is mainly driven by electrostatic interaction, while vdW interaction is pivotal for binding of 29162 (Table S3). To shed more light on the nature of the binding of 9973 and 29162, the atoms of these compounds were classified into four (Fig. 10) and two groups (Fig. 11). The electrostatic and vdW interactions, as well as the number of HBs per snapshot, were calculated for each group separately. For 9973, group 3 contributed dominantly

not only to the electrostatic energy but also to hydrogen bonding (Fig. 11), while the distributions of vdW interactions were relatively equal among all groups. Thus, group 3 plays the key role in the binding affinity of 9973 to CD44.

In the case of 29162, the average number of HBs per snapshot is almost the same for both groups. The differences between them arise from interaction energies. Although group 1 has a great contribution to the vdW interaction, its total contribution is rather modest (approximately $-25.1 \text{ kcal mol}^{-1}$) because of the positive electrostatic energy. On the other hand, having the positive charge of $+0.236 \text{ e}$, the electrostatic interaction of group 2 with negatively charged CD44 (total charge -8e) is attractive, leading to a much stronger binding than group 1 with a total interaction energy of $-57.5 \text{ kcal mol}^{-1}$. Therefore, the binding affinity of 29162 is governed by group 2. Although the vdW interaction drives the stability of CD44 in complex with 29162, the electrostatic energy of group 2 is large. The reason for less contribution of electrostatic interaction is likely the repulsive interaction of group 1 with the receptor.

Table 2 The binding free energies of F-19848A and the top ten SMD hits obtained by MM-PBSA method. The results for F-19848A was averaged over four MD trajectories while one MD run was carried out for the top ten hits. Energy is shown as kcal mol^{-1}

Ligand	ΔE_{elec}	ΔE_{vdw}	ΔG_{PB}	ΔG_{sur}	$-T\Delta S$	ΔG_{bind}
F-19848A	-9.7	-62.2	54.9	-6.5	11.4	-12.1 ± 1.6
9973(9)*	-57.32	-43.60	40.41	-3.81	20.09	-44.23
29162(4)	-27.38	-59.83	37.04	-3.58	14.74	-39.01
31515(10)	-15.99	-51.13	21.62	-4.82	13.42	-36.90
26269(3)	-50.82	-44.68	41.39	-4.33	28.35	-30.09
17528(6)	-11.78	-40.47	17.99	-4.21	9.77	-28.69
29340(5)	-3.23	-42.49	9.6	-3.46	16.59	-22.97
25752(7)	-57.63	-33.38	50.60	-3.29	23.01	-20.69
25497(2)	-75.03	-28.75	62.81	-3.18	23.47	-20.68
25501(1)	-50.86	-43.68	55.62	-3.93	26.04	-16.81
31820(8)	-17.62	-30.49	28.50	-2.90	10.70	-11.81

(*): the number in parentheses indicates the ranking by rupture force of the corresponding ligand

Conclusions

Using computational approaches, we have been able to study in detail the interaction between the F-19848A and CD44 receptor at the atomic level. The calculated binding free energy of this ligand is consistent with other experiments. Application of the virtual screening (docking and SMD) and MM-PBSA method to the TCM database led to the identification of ten ligands which are potentially better or equivalent to F-19848A at binding to CD44. We recommend these ten compounds for further *in vitro* and *in vivo* studies.

Future directions

To further understand CD44/ligand interactions, future work will focus on *in vitro* evaluation of CD44 activity in the

Fig. 10 (a) Groups for compound 9973, a “champion” compound by MM-PBSA simulations. Electrostatic and vdW interaction energies (b) and HBs per snapshot (c) for four groups

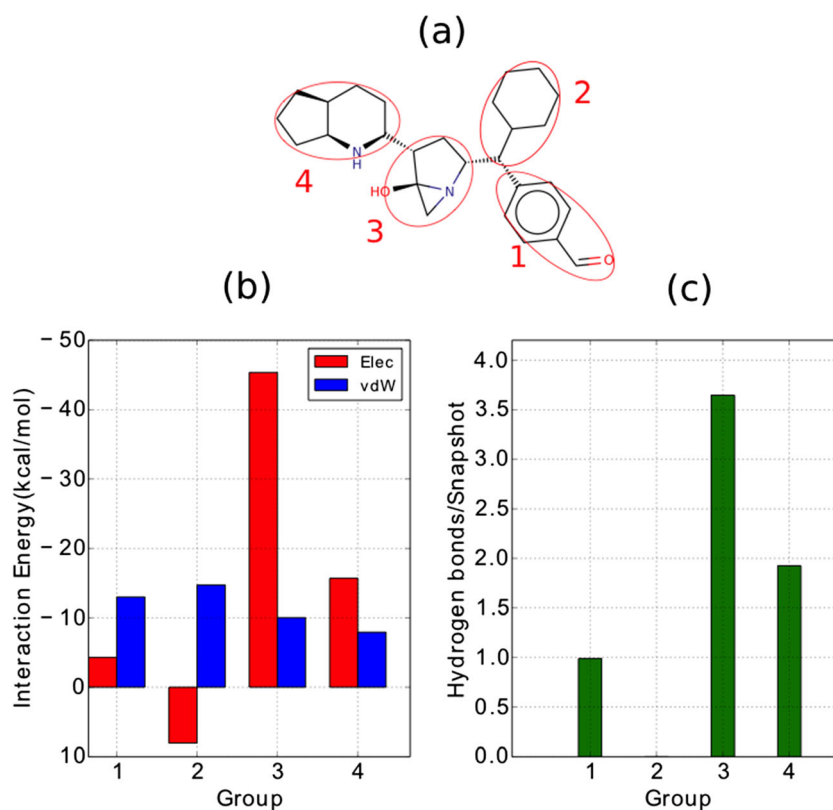
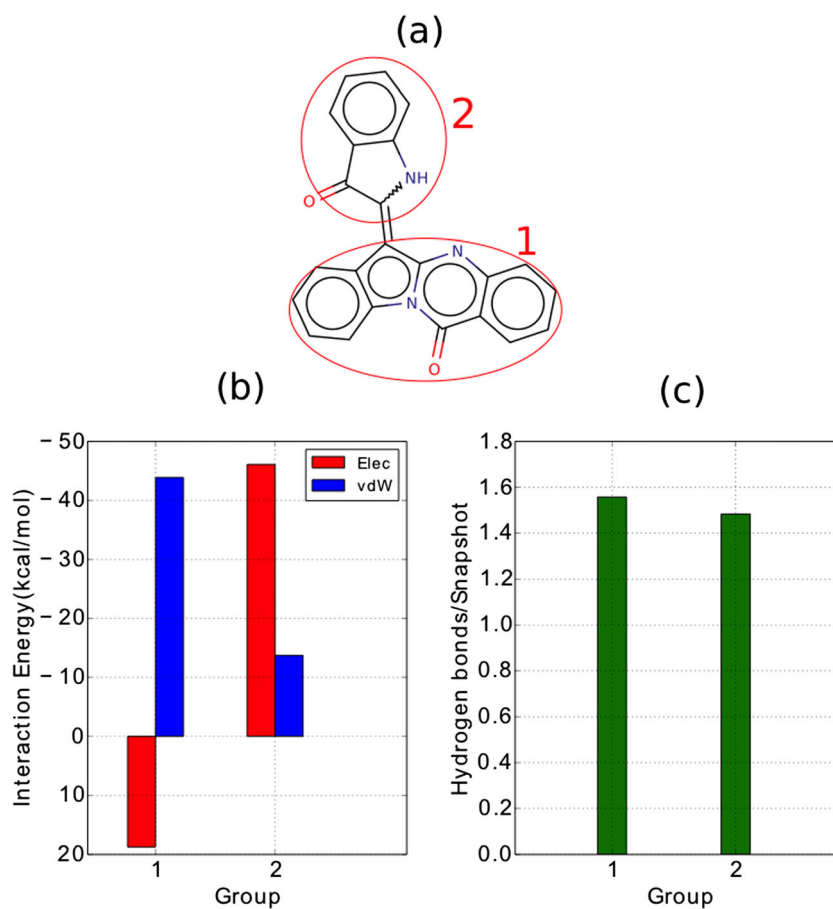


Fig. 11 (a) Groups for compound 29162 which ranked second by MM-PBSA simulations. Electrostatic and vdW interaction energies (b) and HBs per snapshot (c) for two groups



presence of the top ligands predicted in this study to have strong binding to CD44. We plan to expand the search for new potential inhibitors using larger databases as well.

Acknowledgments This work was supported by Department of Science and Technology at Ho Chi Minh city, Vietnam. We are grateful to Quan Van Vuong for useful discussions and technical assistance. Allocation of CPU time at the supercomputer center TASK in Gdansk (Poland) is highly appreciated.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. *CA Cancer J Clin* 55:74–108
- Parkin DM, Fernández LMG (2006) Use of statistics to assess the global burden of breast cancer. *Breast J* 12(Suppl 1):S70–S80
- Guarneri V, Conte PF (2004) The curability of breast cancer and the treatment of advanced disease. *Eur J Nucl Med Mol Imaging* 31(Suppl 1):S149–S161
- Gonzalez-Angulo AM, Morales-Vasquez F, Hortobagyi GN (2007) Overview of resistance to systemic therapy in patients with breast cancer. Springer New York, pp 1–22
- Gokmen-Polar Y, Nakshatri H, Badve S (2011) Biomarkers for breast cancer stem cells: the challenges ahead. *Biomark Med* 5: 661–671
- Iqbal J, Chong PY, Tan PH (2013) Breast cancer stem cells: an update. *J Clin Pathol* 66:485–490
- Al-Hajj M, Wicha MS, Benito-Hernandez A et al (2003) Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 100:3983–3988
- Van Phuc P, Nhan PLC, Nhung TH et al (2011) Downregulation of CD44 reduces doxorubicin resistance of CD44CD24 breast cancer cells. *Oncotargets Ther* 4:71–78
- Ponta H, Sherman L, Herrlich PA (2003) CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol* 4:33–45
- Naor D, Nedvetzki S, Golan I et al (2002) CD44 in cancer. *Crit Rev Clin Lab Sci* 39:527–579
- Guvench O (2015) Revealing the mechanisms of protein disorder and N-glycosylation in CD44-hyaluronan binding using molecular simulation. *Inflammation* 6:305. doi:10.3389/fimmu.2015.00305
- Plazinski W, Knys-Dzieciuch A (2012) Interactions between CD44 protein and hyaluronan: insights from the computational study. *Mol Biosyst* 8:543–547
- Jamison FW, Foster TJ, Barker JA et al (2011) Mechanism of binding site conformational switching in the CD44-hyaluronan protein-carbohydrate binding interaction. *J Mol Biol* 406:631–647
- Favreau AJ, Faller CE, Guvench O (2013) CD44 receptor unfolding enhances binding by freeing basic amino acids to contact carbohydrate ligand. *Biophys J* 105:1217–1226. doi:10.1016/j.bpj.2013.07.041
- Jana M, Bandyopadhyay S (2012) Conformational flexibility of a protein-carbohydrate complex and the structure and ordering of surrounding water. *Phys Chem Chem Phys* 14:6628–6638. doi:10.1039/c2cp24104h
- Jana M, Bandyopadhyay S (2012) Restricted dynamics of water around a protein-carbohydrate complex: computer simulation studies. *J Chem Phys* 137:055102. doi:10.1063/1.4739421
- Hirota-Takahata Y, Harada H, Tanaka I et al (2007) F-19848 A, a novel inhibitor of hyaluronic acid binding to cellular receptor CD44. *J Antibiot* 60:633–639
- Mondal S, Bandyopadhyay S, Ghosh MK et al (2012) Natural products: promising resources for cancer drug discovery. *Anticancer Agents Med Chem* 12:49–75
- Chen CY-C (2011) TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One* 6:e15939–e15939
- Teriete P, Banerji S, Noble M et al (2004) Structure of the regulatory hyaluronan binding domain in the inflammatory leukocyte homing receptor CD44. *Mol Cell* 13:483–496
- Frisch M, Trucks G, Schlegel H et al. (2004) Gaussian 03, revision C.02. Gaussian Inc, Wallingford
- Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461
- Banerji S, Wright AJ, Noble M et al (2007) Structures of the Cd44-hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nat Struct Mol Biol* 14:234–239
- van der Spoel D, Lindahl E, Hess B et al (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718
- Schüttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60:1355–1363
- Van Gunsteren W, Billeter SR, Eising AA, et al. (1996) Biomolecular simulation: the GROMOS96 manual and user guide.
- Berendsen HJC, Postma JPM, Van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. Springer, Dordrecht, pp 331–342
- Van Gunsteren WF, Berendsen HJC (1988) A leap-frog algorithm for stochastic dynamics. *Mol Simul* 1:173–185
- Hess B, Bekker H, Berendsen H, Fraaije J (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18: 1463–1472
- Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089
- Bussi G, Zykova-Timan T, Parrinello M (2009) Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J Chem Phys* 130:074101
- Berendsen HJC, Postma JPM, Van Gunsteren WF et al (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684
- Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52:7182–7190
- Grubmüller H, Heymann B, Tavan P (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science (New York NY)* 271:997–999
- Suan Li M, Khanh Mai B (2012) Steered molecular dynamics-a promising tool for drug design. *Curr Bioinforma* 7:342–351
- Chovancova E, Pavelka A, Benes P et al (2012) CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* 8:e1002708–e1002708
- Gibson CT, Weeks BL, Abell C et al (2003) Calibration of AFM cantilever spring constants. *Ultramicroscopy* 97:113–118
- Mai BK, Viet MH, Li MS (2010) Top leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach. *J Chem Inf Model* 50:2236–2247
- Kollman PA, Massova I, Reyes C et al (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897

40. McQuarrie DA (1973) Statistical thermodynamics, 2nd edn. Harper and Row, New York
41. Baker NA, Sept D, Joseph S et al (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037–10041
42. Duan Y, Wu C, Chowdhury S et al (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012
43. Samsonov SA, Teyra J, Pisabarro MT (2011) Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *J Comput Aided Mol Des* 25:477–489
44. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* 1:337–341
45. Ganesan A (2008) The impact of natural products upon modern drug discovery. *Curr Opin Chem Biol* 12:306–317
46. Vuong QV, Nguyen TT, Li MS (2015) 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* 55:2731–2738
47. Nguyen TT, Mai BK, Li MS (2011) Study of Tamiflu sensitivity to variants of A/H5N1 virus using different force fields. *J Chem Inf Model* 51:2266–2276
48. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 51:69–82
49. Mai BK, Li MS (2011) Neuraminidase inhibitor R-125489—a promising drug for treating influenza virus: steered molecular dynamics approach. *Biochem Biophys Res Commun* 410:688–691
50. Meeprasert A, Rungrotmongkol T, Li MS, Hannongbua S (2014) In silico screening for potent inhibitors against the NS3/4A protease of hepatitis C virus. *Curr Pharm Des* 20:3465–3477
51. Colizzi F, Perozzo R, Scapozza L et al (2010) Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J Am Chem Soc* 132:7361–7371