IDENTIFYING POTENTIAL COMPOUNDS FOR INHIBITION CD44 TARGET OF HUMAN BREAST CANCER STEM CELLS BY DOCKING METHOD

Pham Truong Giang, Tran Quoc Tuan, Huynh Thi Ngoc Thanh, Nguyen Quoc Thai, and Le Thi Ngoc Tu*

Faculty of Natural Sciences Teacher Education, School of Education, Dong Thap University, Cao Lanh 870000, Vietnam

**Corresponding author: Le Thi Ngoc Tu, Email: ltntu@dthu.edu.vn*

Article history

Received: 26/12/2022; Received in revised form: 20/3/2024; Accepted: 22/3/2024

Abstract

In a recent cell-based assay, it has been demonstrated that the F-19848A compound inhibits the binding of hyaluronan to the CD44 receptor as a cell-surface glycoprotein and a receptor for hyaluronan, a major component of the tumor extracellular matrix. The interaction between CD44 and hyaluronan has been shown to promote breast cancer metastasis according to evidence. In this study, the PubChem database contains more than 112 million compounds. The data is inputted for virtual screening to find out top hits by combining Lipinski's rule and docking method. With 20 configurations obtained by docking method, the lowest binding affinity ΔE_{bind} achieved in the best docking mode was chosen as a scoring function for picking out top ligands. For inhibition the CD44 target, the top-leads compounds with binding energy less than -9.0 kcal.mol^{-1} and F-19848A have selected. By docking method, the binding site and other quantities were determined such as the number of hydrogen bonds (HB), non-bond contacts (NBC) of top ligands with CD44 target. Besides, the results also showed that the non-bonded contacts dominate over hydrogen bonds in the interaction between top ligands with CD44 target.

Keywords: *Binding energy, CD44, docking method, Lipinski's rule, small compound.*

DOI: https://doi.org/10.52714/dthu.13.5.2024.1285

Cite: Pham, T. G., Tran, Q. T., Huynh, T. N. T., Nguyen, Q. T., & Le, T. N. T. (2024). Identifying potential compounds for inhibition CD44 target of human breast cancer stem cells by docking method. *Dong Thap University Journal of Science, 13*(5), 30-36. https://doi.org/10.52714/dthu.13.5.2024.1285.

XÁC ĐỊNH NHỮNG HỢP CHẤT TIỀM NĂNG NHẰM ỨC CHẾ THỤ THỂ CD44 CỦA TẾ BÀO GỐC UNG THƯ BẰNG PHƯƠNG PHÁP DOCKING

Phạm Trường Giang, Trần Quốc Tuấn, Huỳnh Thị Ngọc Thanh, Nguyễn Quốc Thái và Lê Thị Ngọc Tú*

Khoa Sư phạm Khoa học Tự nhiên, Trường Sư phạm, Trường Đại học Đồng Tháp, Việt Nam

**Tác giả liên hệ: Lê Thị Ngọc Tú, Email: ltntu@dthu.edu.vn*

Lịch sử bài báo

Ngày nhận: 26/12/2022; Ngày nhận chỉnh sửa: 20/3/2024; Ngày chấp nhận: 22/3/2024

Tóm tắt

CD44 là một glycoprotein xuyên màng và là thụ thể cho Hyaluronan (HA), một trong những thành phần chính của mô ngoại bào khối u. Gần đây, bằng thực nghiệm trên tế bào đã chứng minh rằng hợp chất F-19848A liên kết vào thụ thể CD44 và có khả năng ức chế thụ thể CD44. Trong nghiên cứu này, sử dụng ngân hàng PubChem chứa hơn 112 triệu hợp chất được xem là đầu vào để sàng lọc ảo, tìm ra các hợp chất tiềm năng nhất bằng cách kết hợp quy tắc của Lipinski và phương pháp mô phỏng docking. Kết quả của 20 cấu hình thu được bằng phương pháp docking, với cấu hình tốt nhất có ái lực liên kết thấp nhất ΔE_{bind} được chọn làm hàm tính điểm để chọn ra các phối tử đứng đầu. Để ức chế mục tiêu CD44, các hợp chất đứng đầu phải có năng lượng liên kết nhỏ hơn -9,0 kcal.mol^{-1} và thấp hơn năng lượng liên kết của F-19848A được chọn. Bằng phương pháp mô phỏng docking, vị trí liên kết và các đại lượng khác được nghiên cứu chẳng hạn như số lượng liên kết hydro (HB), tương tác không liên kết (NBC) của các phối tử tiềm năng với mục tiêu CD44. Bên cạnh đó, kết quả cũng chỉ ra rằng các tương tác không liên kết chiếm ưu thế hơn so với liên kết hydro trong tương tác giữa các phối tử đầu với mục tiêu CD44.

Từ khóa: *Hợp chất nhỏ, năng lượng liên kết, phương pháp docking, pro-tê-in CD44, quy tắc Lipinski.*

1. Introduction

The most common cause of cancer-related mortality among women worldwide is breast cancer (Hortobagyi et al., 2005). Despite a notable decrease in overall mortality from breast cancer over the past two decades, currently over 50% of breast tumors do not respond to existing therapies (Gonzalez-Angulo et al., 2007). Thus, there is a pressing need for novel and improved strategies to combat breast cancer. The hypothesis is that breast cancer stem cells (BCSCs) are present in all malignant breast tumors. Accepted ideas are agreed by the scientific community. BCSCs were first identified in human breast tumors by Al Hajj et al. (2003). CD44 is a cell-surface glycoprotein and receptor for hyaluronan, one of the major components of the tumor extracellular matrix (Al-Hajj et al., 2003). The previous researches showed that the inhibition of CD44 suppressed breast tumor growth in mice (Gökmen-Polar et al., 2011). By experiments, many ligands were determined to bind into CD44 at different intracellular and extracellular positions such as hyaluronic acid (HA; also called hyaluronan). The inhibition of HA binding to CD44 could interfere with the turnover and function of CD44, and might be of positive effect in the treatment of many diseases, including osteoarthritis, rheumatoid arthritis and cancer (Naor et al., 2002). However, the understanding of these interactions of these ligands at the molecular level is limited and unclear about the mechanism of the interactions. Like with HA, F-19848A has also been found to bind to the extracellular regions of CD44. The F-19848A molecule can inhibit CD44 activity with half maximal inhibitory concentration (IC50) is 23.5 μM (Berendsen et al., 1981). The experimental binding free energy $\Delta G_{bind}^{exp} = \mathrm{RTln(IC50)}$ was obtained from the IC50 value, where RT = 0.597 kcal/mol at 300K, and IC50 is measured in M. Using this formula, we obtained $\Delta G_{bind}^{exp} = -6.4$ kcal/mol chosen as a scoring function for picking out top ligands from the PubChem database.

A potential therapeutic drug that must require an inhibition constant of nanomolar (nM), meanwhile, the inhibition constant of F-19848A is relatively high (μM), therefore screening to find better candidates to block CD44 activity. For CD44 inhibitors of hyaluronan (HA) binding, in this study, we screened potential lead compounds from the large data base PubChem with 112 million compounds. The combining pharmaceutical characteristics with docking dynamics simulation employed. The results obtained nine (09) potential compounds with binding energy less than -9.0 kcal.mol^{-1} and are potentially better than the F-19848A ligand at binding to CD44 receptor. The binding mechanisms of the top nine hit compounds were evaluated at the atomic level in this study such as hydrogen bonds (HB), non-bond contacts (NBC).

2. Material and methods

2.1. Material

2.1.1. Receptor

The molecular structure of the CD44 and hyaluronic acid (HA) binding domain complex with a small molecule was obtained by experimental method of X-ray crystal diffraction (X-ray crystallography). The structure PDB ID 4NP3 (Liu & Finzel, 2014) was used for simulation with CID 73441667 (HA). The 3D structure of 4NP3 with CID 73441667 indicated in Figure 1.

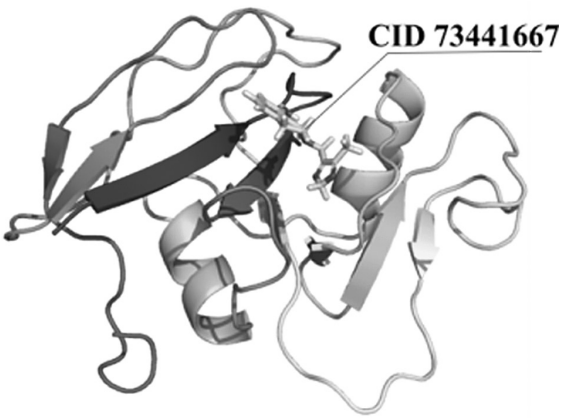

Figure 1. The crystal structure of the CD44 and Hyaluronan (HA) binding domain complex with a small molecule.

2.1.2. Data base of ligands

In 2022, around 112 million chemical structures of compounds were taken from the PubChem database (Bolton et al., 2008), which are investigated as input to the virtual screening process. (https://www.ncbi.nlm.nih.gov/pccompound/limits), up to now (May 2023), which contains 115 million compounds. The

drug-like compounds were filtered by Lipinski's rule (Lipinski et al., 2012) as molecular weight from 0 to 500, XLogP from 0 to 5, hydrogen bond donor count from 0 to 5, hydrogen bond acceptor count from 0 to 10, TPSA from 0 to 140. From the whole set, only 49873 ligands were obtained with the 3D structure for further study.

2.2. Docking method

This study used Autodock Tool 1.5.4 (Sanner, 1999) to prepare PDBQT files for target and ligands. The docking simulation has been carried out by Autodock Vina version 1.1 (Trott & Olson, 2010). Autodock Vina software was used to dock the reduced set of ligands to receptor with the system computer in Division of Physics, School of Education, Dong Thap University, which included 07 high-configuration computer with 11th Gen Intel(R) Core(TM) i9-11900K @ 3.50GHz 3.50 GHz, RAM 32.0 GB, and Centrum Informatyczne TASK, Komputery Duzej Mocy Tryton (see information: https://task.gda.pl/en/resources/superkomputer/computing-power/). To achieve reliable results in global search we set exhaustiveness parameter equal 600. The target ID 4NP3 has binding sites with CID 73441667 known from experiment, the box was chosen to cover just the binding site with grid dimensions 2.0 x 2.5 x 2.0 nm. In docking simulation, the binding site residues of the receptor were flexible. The scoring function is the binding (lowest) energy ΔE_{bind} obtained in the best docking modes.

3. Results and discussion

3.1. Docking scores and best docking poses

With 49873 ligands input, the resulting docking showed that the distributions of docking binding energy ΔE_{bind} of 49873 ligands for the best configuration of Auto Dock Vina is the one with the lowest energy and the RMSD is 0 compared to the others. Those are shown in Figure 2. The binding energies to 4NP3 of CD44 vary from -1.5 kcal.mol^{-1} to -10.9 kcal.mol^{-1}.

From Figure 2 showed that the distributions of binding energies of 49873 ligands to receptor 4NP3 focus mainly with level binding energy within -6.0 kcal.mol^{-1} about 31.0%, while from -9.0 kcal.mol^{-1} to -10 about 0.064% that is frequency of occurrence 0.25. The result of virtual screening by docking simutation obtained 12437 which are potentially better than the F-19848A ligand at binding to CD44 receptor.

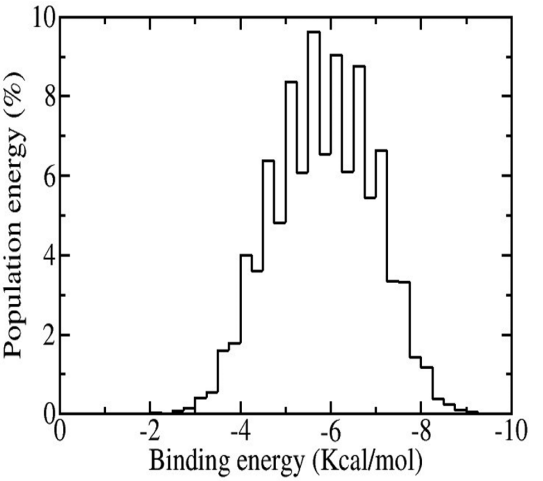

Figure 2. Population of binding energies of 49873 ligands into 4NP3 target. Results were obtained in the best docking mode

Combining conditions (1) the binding energy of ligands into target lower than the binding energy of F-19848A ligand that is control compound, (20) and their binding energy is smaller -9.0 kcal.mol^{-1} (equivalent to nanomolar, nM). From there, 09 compounds were obtained. The locations of these compounds in 4NP3 of CD44 were showed in Figure 3.

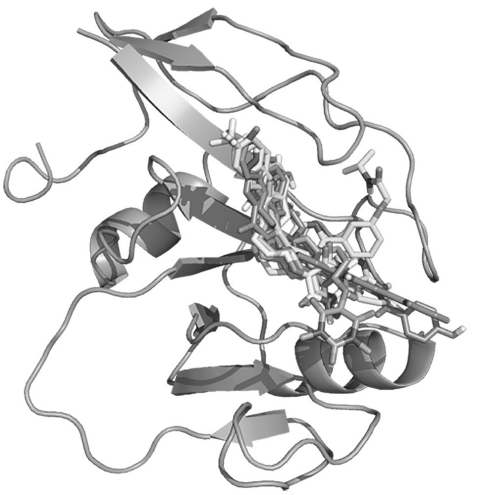

Figure 3. The binding positions of 09 compounds with the binding energy lower than -9.0 kcal.mol^{-1} for 4NSP3. The structures were obtained in the best docking modes

The compounds are inside the binding site of target CD44, which are same HA (CID 73441667). The binding energies of potential compounds with CD44 target showed in Table 1. In the best docking modes with RMSD equivalent to 0, CID 91754535 is the strongest binding into CD44 target, the reason could be the complex structure and position binding of CID 91754535, this issue will be explained in the following section.

Table 1. Nine compounds have the binding energy lower than -9.0 kcal.mol^{-1}

CID	IUPAC name	Lipinski's rule	ΔE_{bind} (kcal.mol^{-1})
11950170	2-[3-[4-(1H-indazol-5-ylamino)quinazolin-2-yl]phenoxy]-N-propan-2-ylacetamide	Molecular Weight: 452.5, XLogP: 4.8 Hydrogen Bond Donor Count: 3 Hydrogen Bond Acceptor Count: 6 Topological Polar Surface Area: 105 Å^2	-9.1
121489294	(3S)-3-(2,4-dihydroxyphenyl)-5-hydroxy-8,8-dimethyl-2,3-dihydropyrano[2,3-h] chromen-4-one	Molecular Weight: 354.4, XLogP: 3.5 Hydrogen Bond Donor Count: 3 Hydrogen Bond Acceptor Count: 6 Topological Polar Surface Area: 96.2 Å^2	-9.0
44340197	(4'aS,11'aS)-8'-(4-fluorophenyl)-1,3,11'a-trimethylspiro[1,3-diazinane-5,3'-4a,5,6,11-tetrahydro-2H-chromeno[5,6-f] indazole]-2,4,6-trione	Molecular Weight: 476.5, XLogP: 2.4 Hydrogen Bond Donor Count: 0 Hydrogen Bond Acceptor Count: 6 Topological Polar Surface Area: 84.7 Å^2	-9.2
44340466	(4'aS,11'aS)-11'a-methyl-8'-phenylspiro [1,3-diazinane-5,3'-4a,5,6,11-tetrahydro-2H-chromeno[5,6-f] indazole]-2,4,6-trione	Molecular Weight: 430.5, XLogP: 2.0 Hydrogen Bond Donor Count: 2 Hydrogen Bond Acceptor Count: 5 Topological Polar Surface Area: 102 Å^2	-9.0
44340182	(4'aS,11'aS)-8'-(4-fluorophenyl)-11'a-methylspiro[1,3-diazinane-5,3'-4a,5,6,11-tetrahydro-2H-chromeno[5,6-f] indazole]-2,4,6-trione	Molecular Weight: 448.4, XLogP: 2.1 Hydrogen Bond Donor Count: 2 Hydrogen Bond Acceptor Count: 6 Topological Polar Surface Area: 102 Å^2	-9.1
3086034	2-[(1,3-dioxoisoindol-2-yl)methyl]-5,12-dihydroquinolino[2,3-b]acridine-7,14-dione	Molecular Weight: 471.5, XLogP: 4.5 Hydrogen Bond Donor Count: 2 Hydrogen Bond Acceptor Count: 6 Topological Polar Surface Area: 95.6 Å^2	-9.0
44340197	(4'aS,11'aS)-8'-(4-fluorophenyl)-1,3,11'a-trimethylspiro[1,3-diazinane-5,3'-4a,5,6,11-tetrahydro-2H-chromeno[5,6-f] indazole]-2,4,6-trione	Molecular Weight: 476.5, XLogP: 2.4 Hydrogen Bond Donor Count: 0 Hydrogen Bond Acceptor Count: 6 Topological Polar Surface Area: 84.7 Å^2	-9.2
89670174	5-[[(1R,1aS,6bR)-1-[6-(trifluoromethyl)-1H-benzimidazol-2-yl]-1a,6b-dihydro-1H-cyclopropa[b][1]benzofuran-5-yl] oxy]-3,4-dihydro-1H-1,8-naphthyridin-2-one	Molecular Weight: 478.4, XLogP: 3.7 Hydrogen Bond Donor Count: 2 Hydrogen Bond Acceptor Count: 8 Topological Polar Surface Area: 89.1 Å^2	-9.0
91754535	4-hydroxy-N-[2-[(1R,13S)-3-methyl-8-oxo-11-azatetracyclo[8.4.0.01,13.02,7]tetradeca-2,4,6,9-tetraene-11-carbonyl]imidazo[1,2-a] pyridin-6-yl]benzamide	Molecular Weight: 490.5, XLogP: 3.9 Hydrogen Bond Donor Count: 2 Hydrogen Bond Acceptor Count: 5 Topological Polar Surface Area: 104 Å^2	-10.0

3.2. The interaction mechanism of the potential compounds and CD44 target

The number of HBs, and NBCs of the potential compounds with CD44 target were computed and obtained in the best docking modes, as shown in Table 2.

Table 2 has showed that the interaction of the potential compounds with CD44 target is mainly the focal residues: Arg155, Asn29, Thr31, among Arg155 has +1e charge. It is suggested that these residues have a significant impact on the interaction between the ligand and the receptor; however, further

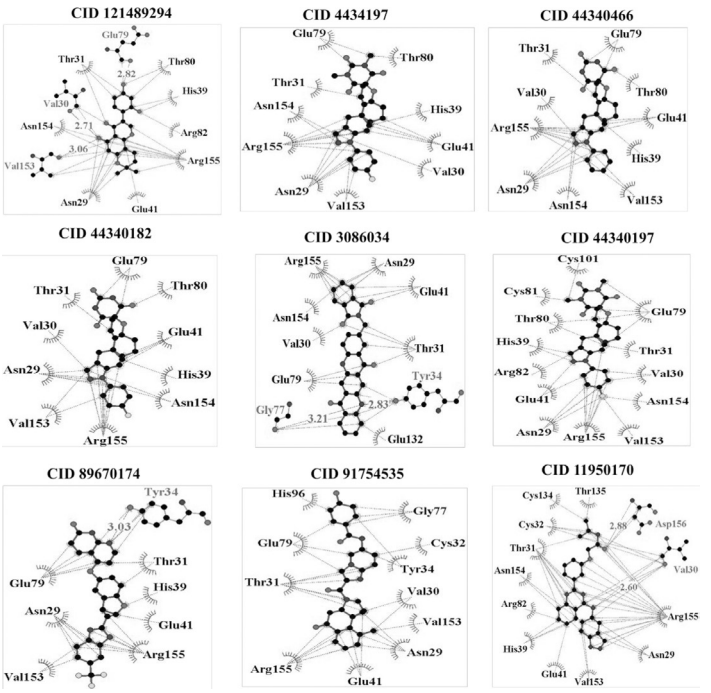

Figure 4. Non-bonded contacts of the potential compounds with CD44 target in the best docking mode. The results are showed by Ligplot version 4.5.3.

clarification through more precise calculations like Molecular Dynamics (MD) is necessary.

From table 1 and table 2, the result found that CID 11950170 its energy affinity is lowest with ΔE_{bind} = -10.0 kcal.mol^{-1} because it has many NBCs (13 NBCs) and 02 HBs. However, CID 121489294 has 03 HBs and 11 NBCs but energy affinity is -9.0 kcal.mol^{-1}. Overall, the binding energies are good correlation with NBCs (Table 1, 2). The results show that the non-bonded contact network is much richer than hydrogen bond network implying that the hydrogen bonding is less important in stabilizing receptor-ligand complexes compared to non-bonded bonds.

The total charge of residues making non-bonded contacts with lead compounds are diversified from -2.0e to 0.0e. For the complex systems of residues with 0.0e, they have many NBCs: CID 11950170 and CID 44340197 have 13 NBCs, CID 121489294 has 11 NBCs, while CID 3086034 with +2e but its NBCs is 10 and for compounds with charge +1.0e or -1.0e make 10-11 NBCs, even is 08 NBCs. Thus, the potential compounds with total charge 0.0e only are strongly interaction with CD44 target for total charge of non-bonded contacts residues, which is 0.0e, but also impact stabilization of receptor-ligand complexes.

Table 2. The number of hydrogen bonds (HBs) and non-bonded contacts (NBCs) of potential compounds with CD44 target.

CID	HBs	NBCs	Charge(s)
11950170	02: Asp15, Val30.	(13): Thr135, Asp156, Val30, **Arg155, Asn29,** Val153, Glu41, His39, Arg82, Asn154, **Thr31,** Cys32, Cys134.	00
121489294	03: Glu79, Val30, Val153.	(11): Thr80, **Arg155,** His39, Arg82, Glu41, **Asn29,** Val153, Asn154, Val30, **Thr31,** Glu79.	00
44340197	00	(10): Asn154, Thr80, His39, Glu41, Val30, Val153, **Asn29, Arg155, Thr31,** Glu79.	-1.0
44340466	00	(10): Asn154, Thr80, Glu41, His39, Val153, **Asn29, Arg155,** Val30, **Thr31,** Glu79.	-1.0
44340182	00	(10): Val153, Thr80, Glu41, His39, Asn154, **Arg155, Asn29,** Val30, **Thr31,** Glu79.	-1.0

3086034	02: Tyr34, Gly77.	(10): Asn154, **Asn29,** Glu41, Tyr34, **Thr31,** Glu132, Gly77, Glu79, Val30, **Arg155.**	-2.0
44340197	00	(13): Val153, Glu79, **Thr31,** Val30, Asn154, **Arg155, Asn29,** Glu41, Arg82, His39, Thr80, Cys81, Cys101.	0.0
89670174	01: Tyr34	(08): **Arg155,** Val30, **Thr31,** His39, Glu41, Val153, **Asn29,** Glu79.	-1.0
91754535	00	(11): Val153, Gly77, Cys32, Tyr34, Val30, **Asn29,** Glu41, **Arg155, Thr31,** Glu79, His96.	-1.0

4. Conclusions

By combining Lipinski's rule, docking method, and experiment, we obtained the binding energies of 09 potential compounds for inhibition CD44 target with binding energy less than -9.0 kcal.mol^{-1} as an inhibition constant of nanomolar (nM) and these binding energies are potentially better than the F-19848A ligand (control compound) at binding to CD44 receptor with binding energy -6.4 kcal.mol^{-1}. The results show that the non-bonded contacts dominate over hydrogen bonds in the interaction between the potential compounds with CD44 target. Furthermore, the neutral residues of CD44 target are only important to the non-bonded contacts but also impact stabilization of receptor-ligand complexes. However, this interaction mechanism needs to be re-evaluated with more precise methods (redock) such as steered molecular dynamics (SMD), molecular dynamics (MD) because in docking simulation the receptor dynamics was omitted, therefore the results are less accurate. Yet, the docking method has been evaluated by the researches. It is considered to be a very effective tool in the virtual screening of large compound data. Autodock Tool has 3600 citations; Meanwhile, AutoDock Vina has 23891 citations.

Acknowledgment: This research is supported by the project SPD2022.02.05.

References

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, *100*(7), 3983-3988.

Berendsen, H. J., Postma, J. P., van Gunsteren, W. F., & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In *Intermolecular forces* (pp. 331-342). Springer.

Bolton, E. E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). PubChem: integrated platform of small molecules and biological activities. *Annu Rep Comput Chem*, *4*, 217-241.

Gökmen-Polar, Y., Nakshatri, H., & Badve, S. (2011). Biomarkers for breast cancer stem cells: the challenges ahead. *Biomarkers in medicine*, *5*(5), 661-671.

Hortobagyi, G. N., de la Garza Salazar, J., Pritchard, K., Amadori, D., Haidinger, R., Hudis, C. A., . . . Namer, M. (2005). The global breast cancer burden: variations in epidemiology and survival. *Clinical breast cancer*, *6*(5), 391-401.

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, *64*, 4-17.

Liu, L.-K., & Finzel, B. C. (2014). Fragment-based identification of an inducible binding site on cell surface receptor CD44 for the design of protein–carbohydrate interaction inhibitors. *Journal of medicinal chemistry*, *57*(6), 2714-2725.

Naor, D., Nedvetzki, S., Golan, I., Melnik, L., & Faitelson, Y. (2002). CD44 in cancer. *Critical reviews in clinical laboratory sciences*, *39*(6), 527-579.

Sanner, M. F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model.*, *17*(1), 57-61.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, *31*(2), 455-461.