12-23-2022

In silico activity and ADMET profiling of phytochemicals from Ethiopian indigenous aloes using pharmacophore models

Lemessa Etana Bultum

Gemechu Bekele Tolossa

Gwangmin Kim

Ohhyeon Kwon

Doheon Lee

scientific reports

Check for updates

OPEN

In silico activity and ADMET profiling of phytochemicals from Ethiopian indigenous aloes using pharmacophore models

Lemessa Etana Bultum[1,2,3], Gemechu Bekele Tolossa[2,4], Gwangmin Kim[1,2], Ohhyeon Kwon[1,2] & Doheon Lee[1,2✉]

In silico profiling is used in identification of active compounds and guide rational use of traditional medicines. Previous studies on Ethiopian indigenous aloes focused on documentation of phytochemical compositions and traditional uses. In this study, ADMET and drug-likeness properties of phytochemicals from Ethiopian indigenous aloes were evaluated, and pharmacophore-based profiling was done using Discovery Studio to predict therapeutic targets. The targets were examined using KEGG pathway, gene ontology and network analysis. Using random-walk with restart algorithm, network propagation was performed in CODA network to find diseases associated with the targets. As a result, 82 human targets were predicted and found to be involved in several molecular functions and biological processes. The targets also were linked to various cancers and diseases of immune system, metabolism, neurological system, musculoskeletal system, digestive system, hematologic, infectious, mouth and dental, and congenital disorder of metabolism. 207 KEGG pathways were enriched with the targets, and the main pathways were metabolism of steroid hormone biosynthesis, lipid and atherosclerosis, chemical carcinogenesis, and pathways in cancer. In conclusion, in silico target fishing and network analysis revealed therapeutic activities of the phytochemicals, demonstrating that Ethiopian indigenous aloes exhibit polypharmacology effects on numerous genes and signaling pathways linked to many diseases.

Healthcare systems around the world are encountering increased levels of pandemics, chronic illnesses, population aging and escalating healthcare costs. Patients and healthcare providers alike are demanding that healthcare services be revitalized. This includes improving access to traditional, complementary, and alternative medicine (TCAM) practitioners, products, and services, particularly in primary healthcare[1]. Natural products' bioactive components have been widely used for treating a broad spectrum of diseases for centuries and are considered as promising alternative therapeutic agents due to their potential healing effects[2]. Over the past three decades, there has been a significant increase in the usage of natural product medicines, with 80% of people globally relying on them for some aspect of basic healthcare[3]. The use of plants as a source of natural products and as the basis for traditional medicine systems across many different cultures has been extensively documented for thousands of years[2,4,5].

Plants of the genus Aloe are the most common traditional medicinal plants in recorded history [6]. According to literature records, 74% of the uses of the genus Aloe were as medicine, and exudate was frequently used in these applications[7]. Aloe leaves have been used for thousands of years as a primary remedy for skin conditions, burns, and constipation[8]. Additionally, there are numerous commercial products that contain aloe species, such as laxative medications, health beverages and tonics, after-shave gel, mouthwash and toothpaste, hair tonic and shampoo, and skin-moisturizing gels[9,10]. Aloes were important in Ethiopia for their therapeutic properties as well as for securing livelihoods, fostering economic growth, and protecting biodiversity[9]. With 46 known aloe species and three subspecies, Ethiopia is one of the likely centers of aloe biodiversity. Of these species, 67.3% are indigenous to the nation[11,12]. In Ethiopia, three regional centers of endemism have been identified[10]: (1) North

[1]Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291Daehak-Ro, Daejeon 34141, South Korea. [2]Bio-Synergy Research Center, 291Daehak-Ro, Daejeon 34141, South Korea. [3]Department of Applied Bioscience, Dong-A Universtiy, Busan 49315, South Korea. [4]Department of Neuroscience, Washington University School of Medicine, St. Louis, MO 63110, USA. ✉email: dhlee@kaist.ac.kr

and central highlands and west of the great rift valley have endemic species and infraspecific taxa of *Aloe adigratana*, *A. ankoberensis*, *A. benishangulana*, *A. camperi*, *A. clarkei*, *A. debrana*, *A. elegans*, *A. monticola*, *A. percrassa*, *A. pulcherrima*, *A. schelpei*, *A. sinana*, *A. steudneri*, *A. trigonantha*, and *A. weloensis*; (2) Eastern highlands and lowlands contains endemic species of *Aloe bertemariae*, *A. harlana*, *A. mcloughlinii*, *A. megalacantha* subsp. alticola, *A. pirottae*; *A. pubescens*; and *A. trichosantha* subsp. Longiflora; and (3) Southern highlands, lowlands and rift valley with species *Aloe elkerriana*, *A. friisii*, *A. ghibensis*, *A. gilbertii* subsp. gilbertii, *A. gilbertii* subsp. megalacanthoides, *A. jacksonii*, *A. kefaensis*, *A. otallensis*, *A. tewoldei*, *A. welmelensis*, and *A. yavellana*.

A great deal of effort is being done to identify the aloes' phytochemical compositions because of their extensive medicinal uses for long period of time. Aloe species' leaves and roots are reservoirs for a variety of intriguing secondary metabolites, including alkaloids, anthraquinones, pre-anthraquinones, anthrones, bianthraquinoids, chromones, flavonoids, coumarins, and pyrones[8]. However, major active ingredients, associated diseases, and their mode of action, particularly for indigenous Ethiopian aloes, are not well understood. This hampered their widespread usage in the development of drugs.

Finding bioactive compounds to aid in the treatment of diseases is the goal of the drug development process. In the past, the new drug discovery process begins with random screening and empirical observations of the effects of natural products for known diseases. Currently, this process is improved by high throughput screening (HTS), which enabled the screening process of many thousands of compounds against a molecular target or cellular assay very quickly[13]. Furthermore, researchers are constantly investigating new methods to increase the efficiency of the drug discovery process[14]. Utilizing computer-aided drug design (CADD), often known as the in silico method, is one strategy to improve the effectiveness of novel drug development. Through the use of molecular modeling that enables virtually screening a large number of compounds for drug-likeness properties and their interaction with pharmacological targets, CADD reduces of the drug development time and overall research expenses[13]. Compound in silico activity profiling is an emerging computational method for predicting the most likely targets of a bioactive compound and therefore anticipating adverse reactions, side effects, and drug repurposing[15]. It can be applied in many different fields including drug repositioning, natural product profiling and multitarget approaches[16]. Studies on the new uses of already established drugs are being carried out in addition to classical drug development. In silico profiling of old drugs can help to reposition such old drugs for new uses[17,18]. Compounds that emerge from phenotypic-based screening including empirical knowledge have usually a completely unknown mechanism of action (MoA). Pharmacophore-based activity profiling of natural products enables to predict their mechanism of action and guides in proper planning of in vivo and in vitro experiments. In addition, previously unknown lead therapeutic natural products with interesting chemical scaffolds can be profiled[16]. This will help to prioritize compounds based on their predicted target profile. The understanding and the prediction of the synergistic role of polypharmacology is also one of the main applications of in silico target profiling[19].

Pharmacophore model-based in silico profiling for the compounds from the Ethiopian indigenous aloes was done in this study. A pharmacophore is an abstract description of the molecular features (both steric and electronic) required for the molecular recognition of ligands by biological macromolecules, influencing their biological processes[20]. Structurally distinct molecules with similar pharmacophore patterns are likely to be recognized by the same binding site of a biological target and thus have similar biological profiles[21]. Pharmacophore models are typically used when several drugs (active compounds) have been identified, but the three-dimensional (3D) structure of the target protein or receptor is unknown. Drugs are superimposed to determine their common properties, providing a pharmacophore model that describes ligand-receptor binding[22]. Any atom or group in a molecule with specific properties related to molecular recognition can be reduced to pharmacophore properties[23]. These molecular patterns could be hydrogen bond donors or acceptors, cationic, anionic, aromatic or hydrophobic, and any possible combination[24]. For a set of active molecules, pharmacophore identification involves two steps: analyzing the molecule to identify its pharmacophore properties, that is, atoms that can interact with a receptor, and aligning the active conformations of the molecules to find the best overlay of the corresponding features[22].

In this study, ADMET profiling and pharmacophore model-based in silico target prediction of the phytochemicals from Ethiopian indigenous aloes was performed. To identify related diseases and reveal the mechanism underlying the therapeutic effects of aloes' phytochemicals, gene ontology (GO), KEGG pathway[25], and network analysis of the predicted targets were investigated. To the best of our knowledge, no computerized comprehensive target fishing approach has been used to identify the pharmacological potential of Ethiopian indigenous aloes' phytochemicals.

Results and discussion
Evaluation of physicochemical, drug-likeness and ADMET related properties of phytochemicals from Ethiopian indigenous aloes.

Physicochemical properties affect pharmacodynamics and pharmacokinetics of molecules. Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties of a molecule are important factors in successful drug and other industrial chemicals development[26,27]. Utilizing physicochemical properties that were compiled from the PubChem database[28] and SwissADME webserver[29], the phytochemicals' drug-likeness was assessed. ADMET-related properties of phytochemicals from Ethiopian indigenous aloes were predicted using admetSAR web server—a database which contains ADMET-associated properties curated form a large number of diverse literatures[27]. Prior to in silico profiling of the phytochemicals, we excluded potentially harmful substances that are AMES mutagens, carcinogens, and hERG inhibitors using the data from admetSAR.

We used Lipinski's rule of 5[30] and Veber's rule[31] to assess the drug-likeness properties of phytochemicals from Ethiopian indigenous aloes. However, it should be noted that many of the most successful drugs do not fit these

No	Compound	Molecular weight (g/mol)	Lipophilicity (log p)	Hydrogen bond donors	Hydrogen bond acceptors	TPSA (Å^2)	ROTB	Number of Lipinsk's rules violations	Number of Veber's rules violations
1	Aloe-emodin	270.24	1.8	3	5	94.8	1	0	0
2	Homonataloin	432.4	1.1	6	9	157	3	0	1
3	Microdontin A	564.5	1.6	7	11	194	7	3	1
4	Microdontin B	550.6	1.6	7	10	177	7	2	1
5	Benzyl salicylate	228.24	3.2	1	3	46.5	4	0	0
6	Diisobutyl phtha-late	278.34	4.1	0	4	52.6	8	0	0
7	Ethyl salicylate	166.17	3	1	3	46.5	3	0	0
8	Thiophen-2-yl benzoate	204.25	3.3	0	3	54.5	3	0	0
9	1-Hexadecanol	242.44	7.3	1	1	20.2	14	1	1
10	cis-Linoleic acid	266.4	6	1	2	37.3	13	1	1
11	Methyl 9-methyltet-radecanoate	256.42	6.6	0	2	26.3	13	1	1
12	Methyl decanoate	186.29	4.7	0	2	26.3	9	0	0
13	Methyl dode-canoate	214.34	5.8	0	2	26.3	11	0	1
14	Methyl linoleate	294.5	6.9	0	2	26.3	15	1	1
15	Methyl palmitate	270.5	7.9	0	2	26.3	15	1	1
16	Methyl tetrade-canoate	242.4	6.8	0	2	26.3	13	0	1
17	Palmitic acid	256.42	6.4	1	2	37.3	14	1	1
18	Undecylenic acid	184.27	3.9	1	2	37.3	9	0	0
19	7-O-Methylaloesin	408.4	−1	4	9	143	5	0	1
20	Aloenin	410.4	0	5	10	155	5	0	1
21	Orcinyl angelate	206.24	2.9	1	3	46.5	3	0	0
22	(E)-2-Isopropyl-5-methylphenyl 2-methylbut-2-enoate	232.32	4.2	0	2	26.3	4	0	0
23	(R)-Ar-Turmerone	216.32	4	0	1	17.1	4	0	0
24	Ar-Turmerone	216.32	4	0	1	17.1	4	0	0
25	Carvacrol	150.22	3.1	1	1	20.2	1	0	0
26	Curlone	218.33	4	0	1	17.1	4	0	0
27	Diosgenin,dehydro	396.6	6.9	0	2	18.5	0	1	0
28	Hinesol	222.37	3.7	1	1	20.2	1	0	0
29	Nerol	154.25	2.9	1	1	20.2	4	0	0
30	Phytol	296.5	8.2	1	1	20.2	13	1	1
31	Turmerone	218.33	3.3	0	1	17.1	4	0	0
32	Heptadecane	240.5	8.8	0	0	0	14	1	1
33	Pentadecane	212.41	7.7	0	0	0	12	1	1
34	beta-Sitosterol	414.7	9.3	1	1	20.2	6	1	0
35	1-Heptadecene	238.5	9.5	0	0	0	14	1	1

Table 1. Lipinski's rule of five and Veber's rule for drug-likeness analysis of Ethiopian indigenous aloes' phytochemicals. *TPSA* topological polar surface area, *ROTB* number of rotatable bonds.

guidelines[32]. For oral drugs from natural products or natural product derivatives, 2 or 3 rule of five violations are common[33]. As shown in Table 1, the majority of phytochemicals comply with Veber's rule criteria and the rule of five (with a maximum of 3 rule violations). Thus, the phytochemicals found in indigenous Ethiopian aloes demonstrated adequate drug-likeness and physicochemical characteristics indicating their potential as therapeutic medication.

ADME-related properties such as molecular weight (MW), octanol–water partition coefficients (Log P), topological polar surface area (TPSA), water solubility, gastrointestinal absorption (GIA), and blood–brain bar-rier (BBB) permeability were presented in Table 2. Drug candidates exhibit poor absorption when their TPSA is higher than 140 Å^2, which is benchmarked for marketed drugs[31,34,35]. TPSA has a positive correlation with mass and the molecules with a mass higher than 500 g/mol were observed to have TPSA beyond the range of 0–140[35]. TPSA values and GIA for the majority of the phytochemicals from indigenous Ethiopian aloes were acceptable. The Log p value and Abbott bioavailability score of greater than zero indicate that the phytochemicals have a

No	Compound	Molecular formula	Molecular weight (g/mol)	Water solubility Silicos-IT class	GI absorption	BBB permeant	Abbott bioavailability score
1	Aloe-emodin	$C_{15}H_{10}O_5$	270.24	Soluble	High	No	0.55
2	Homonataloin	$C_{22}H_{24}O_9$	432.4	Soluble	High	No	0.55
3	Microdontin A	$C_{30}H_{28}O_{11}$	564.5	Soluble	Low	No	0.17
4	Microdontin B	$C_{30}H_{30}O_{10}$	550.6	Moderately soluble	Low	No	0.17
5	Benzyl salicylate	$C_{14}H_{12}O_3$	228.24	Moderately soluble	High	Yes	0.55
6	Diisobutyl phthalate	$C_{16}H_{22}O_4$	278.34	Moderately soluble	High	Yes	0.55
7	Ethyl salicylate	$C_9H_{10}O_3$	166.17	Soluble	High	Yes	0.55
8	Thiophen-2-yl benzoate	$C_{11}H_8O_2S$	204.25	Soluble	High	Yes	0.55
9	1-Hexadecanol	$C_{16}H_{34}O$	242.44	Moderately soluble	High	Yes	0.55
10	cis-Linoleic acid	$C_{17}H_{30}O_2$	266.4	Moderately soluble	High	Yes	0.85
11	Methyl 9-methyltetrade-canoate	$C_{16}H_{32}O_2$	256.42	Moderately soluble	High	Yes	0.55
12	Methyl decanoate	$C_{11}H_{22}O_2$	186.29	Soluble	High	Yes	0.55
13	Methyl dodecanoate	$C_{13}H_{26}O_2$	214.34	Moderately soluble	High	Yes	0.55
14	Methyl linoleate	$C_{19}H_{34}O_2$	294.5	Moderately soluble	High	No	0.55
15	Methyl palmitate	$C_{17}H_{34}O_2$	270.5	Poorly soluble	High	Yes	0.55
16	Methyl tetradecanoate	$C_{15}H_{30}O_2$	242.4	Moderately soluble	High	Yes	0.55
17	Palmitic acid	$C_{16}H_{32}O_2$	256.42	Moderately soluble	High	Yes	0.85
18	Undecylenic acid	$C_{11}H_{20}O_2$	184.27	Soluble	High	Yes	0.85
19	7-O-Methylaloesin	$C_{20}H_{24}O_9$	408.4	Soluble	Low	No	0.55
20	Aloenin	$C_{19}H_{22}O_{10}$	410.4	Soluble	Low	No	0.55
21	Orcinyl angelate	$C_{12}H_{14}O_2$	206.24	Soluble	High	Yes	0.55
22	(E)-2-Isopropyl-5-meth-ylphenyl 2-methylbut-2-enoate	$C_{15}H_{20}O_2$	232.32	Moderately soluble	High	Yes	0.55
23	(R)-Ar-Turmerone	$C_{15}H_{20}O$	216.32	Moderately soluble	High	Yes	0.55
24	Ar-Turmerone	$C_{15}H_{20}O$	216.32	Moderately soluble	High	Yes	0.55
25	Carvacrol	$C_{10}H_{14}O$	150.22	Soluble	High	Yes	0.55
26	Curlone	$C_{15}H_{22}O$	218.33	Soluble	High	Yes	0.55
27	Diosgenin,dehydro	$C_{27}H_{40}O_2$	396.6	Moderately soluble	Low	No	0.55
28	Hinesol	$C_{15}H_{26}O$	222.37	Soluble	High	Yes	0.55
29	Nerol	$C_{10}H_{18}O$	154.25	Soluble	High	Yes	0.55
30	Phytol	$C_{20}H_{40}O$	296.5	Moderately soluble	Low	No	0.55
31	Turmerone	$C_{15}H_{22}O$	218.33	Soluble	High	Yes	0.55
32	Heptadecane	$C_{17}H_{36}$	240.5	Poorly soluble	Low	No	0.55
33	Pentadecane	$C_{15}H_{32}$	212.41	Moderately soluble	Low	No	0.55
34	beta-Sitosterol	$C_{39}H_{50}O$	414.7	Poorly soluble	Low	No	0.55
35	1-Heptadecene	$C_{17}H_{34}$	238.5	Poorly soluble	Low	No	0.55

Table 2. Calculated physicochemical properties of Ethiopian indigenous aloes' phytochemicals.

substantial bioavailability and cross the cell membrane efficiently[36]. GIA and BBB permeability are important properties of a drug that is intended for widespread use. There have been numerous prior researches to improve GI permeation of molecules for orally delivered, poorly absorbed medicines[37]. A significant majority of the phytochemicals from indigenous Ethiopian aloes exhibited acceptable GIA and cross the blood–brain barrier.

Permeability glycoprotein (P-gp) influences the ADMET properties of many xenobiotics (foreign drugs or chemicals), and it is important to investigate the interaction between the transporter protein and the drug molecule[38]. It limits cellular uptake and metabolism of compounds by acting as a unidirectional efflux pump to extrude its substrate from inside to outside of cells[39,40]. From Table 3, it can be seen that all of the phytochemicals except phytol were not substrates of P-glycoprotein, demonstrating their desirable properties as potential therapeutic agents.

Cytochrome P450s (CYPs) represent a large class of heme-containing enzymes that are necessary for the detoxification of foreign chemicals and the metabolism of drugs. Cytochrome P450 enzymes can be inhibited or induced by drugs, resulting in clinically significant drug-drug interactions that can cause unanticipated adverse reactions or therapeutic failures. Inhibitors block the metabolic activity of one or more CYP450 enzymes whereas inducers increase CYP450 enzyme activity by increasing enzyme synthesis. Hence, cautions should be taken when prescribing a drug known to be a CYP450 inhibitor or inducer. The target drug may need to be substituted or the dose adjusted to account for a potential decrease or increase in metabolism[41].

No	Compound	Pgp substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor
1	Aloe-emodin	No	Yes	No	No	No	Yes
2	Homonataloin	No	Yes	No	No	No	Yes
3	Microdontin A	No	No	No	No	No	No
4	Microdontin B	No	No	No	No	No	No
5	Benzyl salicylate	No	Yes	Yes	No	No	No
6	Diisobutyl phthalate	No	Yes	Yes	No	No	No
7	Ethyl salicylate	No	No	No	No	No	No
8	Thiophen-2-yl benzoate	No	Yes	Yes	Yes	No	No
9	1-Hexadecanol	No	Yes	No	No	No	No
10	cis-Linoleic acid	No	Yes	No	Yes	No	No
11	Methyl 9-methyltetradecanoate	No	No	No	Yes	No	No
12	Methyl decanoate	No	No	No	No	No	No
13	Methyl dodecanoate	No	Yes	No	No	No	No
14	Methyl linoleate	No	Yes	No	Yes	No	No
15	Methyl palmitate	No	Yes	No	No	No	No
16	Methyl tetradecanoate	No	Yes	No	No	No	No
17	Palmitic acid	No	Yes	No	Yes	No	No
18	Undecylenic acid	No	No	No	No	No	No
19	7-O-Methylaloesin	No	No	No	No	No	No
20	Aloenin	No	No	No	No	No	No
21	Orcinyl angelate	No	No	No	No	No	No
22	(E)-2-Isopropyl-5-methylphenyl 2-methylbut-2-enoate	No	No	No	No	No	No
23	(R)-Ar-Turmerone	No	No	No	No	No	No
24	Ar-Turmerone	No	No	No	No	No	No
25	Carvacrol	No	Yes	No	No	No	No
26	Curlone	No	No	Yes	Yes	No	No
27	Diosgenin,dehydro	No	No	No	Yes	No	No
28	Hinesol	No	No	No	Yes	No	No
29	Nerol	No	No	No	No	No	No
30	Phytol	Yes	No	No	Yes	No	No
31	Turmerone	No	No	No	Yes	No	No
32	Heptadecane	No	Yes	No	No	No	No
33	Pentadecane	No	Yes	No	No	No	No
34	beta-Sitosterol	No	No	No	No	No	No
35	1-Heptadecene	No	Yes	No	No	No	No

Table 3. Interaction of Ethiopian indigenous aloes' phytochemicals with P-glycoprotein and cytochrome P450 isoenzymes.

Pharmacophore model-based in silico profiling of phytochemicals from Ethiopian indigenous aloes. The profiling results were shown in two HTML tables designated MoleculeFits and Pharmacophore-Fits. The two measures utilized to assess the ligand and pharmacophore's fitness were fit-value and shape similarity. Targets were chosen from the ligand profiler using a fit-value criterion of equal to or higher than 0.5.

Finally, 458 pharmacophore models were mapped. The models belonged to 272 target proteins/genes with unique UniProt-ID. 82 of the targets were related to human, and the CODA[42] network was employed for further analysis. A complete list of the 1,885 phytochemical-target pairs with target short name, gene and protein IDs, UniProt-ID is presented in (Supplementary Table S1 online).

Inferring the pharmacological actions using KEGG, GO and network analysis. In this work, a total of 35 compounds found in 10 aloes were successfully mapped to 82 human related targets using pharmacophore models. 207 KEGG pathways[25] were enriched with the 82 targets of Ethiopian indigenous aloes' phytochemicals (Supplementary Table S2 online). Based on the adjusted P-value of less than or equal to 0.05, 104 KEGG pathways were filtered. Figure 1a showed top 20 significantly enriched KEGG pathways. A drug-target-pathway network constructed using python NetworkX library version 2.5 illustrated that 47 genes remarkably associated with the top 20 KEGG pathways. Most of the candidates from the network were involved in seven major KEGG pathways, namely metabolism of steroid hormone biosynthesis, lipid and atherosclerosis, chemical carcinogenesis, IL-17 signaling pathway, Th17 cell differentiation and pathways in cancer pathways (Fig. 1b). KEGG enrichment analysis highlighted targets and pathways where the phytochemicals could effectively exe-
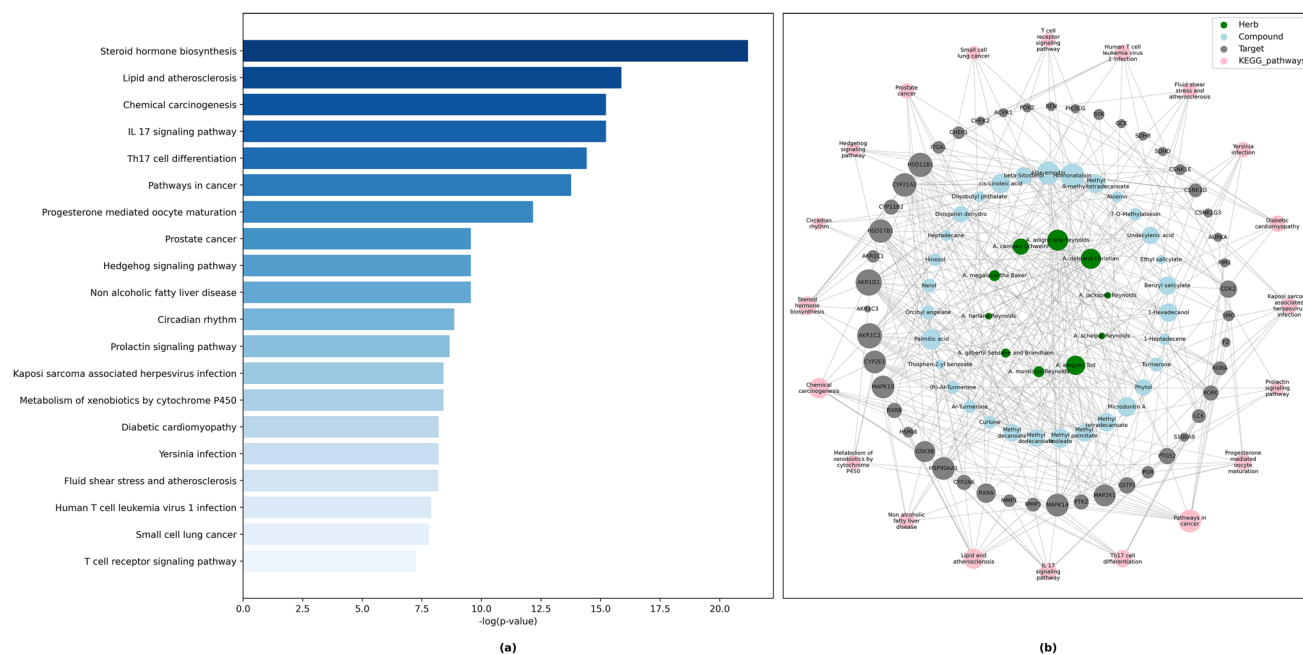
Figure 1. (**a**) Bar graph of KEGG analysis for top 20 enriched pathways. (**b**) Phytochemical-target-pathway network constructed by python NetworkX library illustrated pathways and their associated genes (Adjusted P-value ≤ 0.05).

cute their therapeutic potential. Steroid hormones play an essential role in regulating water and salt balance, metabolism and stress response, and in initiating and maintaining sexual differentiation and reproduction[43]. Previous studies show that aloes can be used as a biological vehicle for delivery of the hormone[44]. Evidence have shown that inflammation plays an essential role in the development, progression, and outcome of atherosclerosis. Owing to their anti-inflammatory properties, aloes may be a good candidate to prevent the progression of atherosclerosis[45]. The phytochemicals found in aloes, especially aloe-emodin, have multiple anti-proliferative and anti-carcinogenic properties in various human cancer cell lines, often affecting multiple important signaling pathways. The most notable effects are inhibition of cell proliferation, cell migration and invasion, cell cycle arrest, induction of cell death, mitochondrial membrane and redox disturbances, and regulation of immune signaling. Aloe-emodin effects are not ubiquitous in all cell lines and vary by cell type[46,47]. Whether looking at aloe as a whole or looking at its compounds, there are various articles that aloes are medicinal plants that have been shown to be beneficial against many types of cancers[47]. However, the precise mechanism of action of aloes' phytochemicals on underlying molecular targets remains to be investigated.

GO analysis[48] is a useful tool for discovering biological processes, cellular components and molecular functions of genes[49]. A total of 1376 biological processes (BPs), 213 molecular functions (MFs), and 101 cellular components (CCs) related to the 82 predicted target proteins/genes were identified (Supplementary Table S2 online). Based on the *P*-value of less than or equal to 0.05, 211 BPs, 58 MFs and 9 CCs were filtered. The functional enrichment analysis of the top 20 BPs demonstrated that most of the targets were correlated with the protein phosphorylation and modification processes, steroid metabolic processes and so on, as shown in Fig. 2a. In addition, a drug-target-biological process network constructed using python NetworkX library version 2.5 illustrated that 55 targets were predominantly enriched with the top 20 BPs (Fig. 2b).

The top 20 functionally enriched MFs are presented in Fig. 3. Among the major 20 MFs, protein serine/threonin kinase activity and ketosteroid monooxygenase activity are dominantly expressed by the 55 target genes (Fig. 3a). Figure 3b shows a drug-target-biological process network illustrating that 45 genes were predominantly enriched with the top 20 MFs.

Furthermore, CCs are widely distributed in the lumen of cytoplasmic vesicles, ficolin-1-rich granules, secretory granules and so on (Fig. 4a and b).

In addition, we performed KEGG pathways, GO and network analysis for each of the 10 aloe species used in this study. The figures are presented in Supplementary Figure S1 online. 15 genes related to *A. adigratana* Reynolds' phytochemicals were enriched in four major KEGG pathways, namely steroid hormone biosynthesis, chemical carcinogenesis, metabolism of xenobiotics by cytochrome p450 and natural killer cell mediated cytotoxicity. The functional enrichment analysis reveals that most of the targets of *A. adigratana* Reynolds' phytochemicals were related to steroid metabolic and biosynthetic process, progesterone metabolic process, glucocorticoid metabolic and biosynthetic process and positive regulation of reactive oxygen metabolic processes. Moreover, ketosteroid monooxygenase activity, steroid dehydrogenase activity, alditol:NADP + 1-oxidoreductase activity, oxidoreductase activity, and alcohol dehydrogenase (NADP +) activity were the major MFs related to *A. adigratana* Reynolds' phytochemicals. Similarly, 23 targets were predicted for phytochemicals found in *A. camperi* Schweinf. These targets were enriched in KEGG pathways of steroid hormone biosynthesis, chemical carcinogenesis, Th17 cell differentiation, circadian rhythm, progesterone-mediated oocyte maturation, cellular
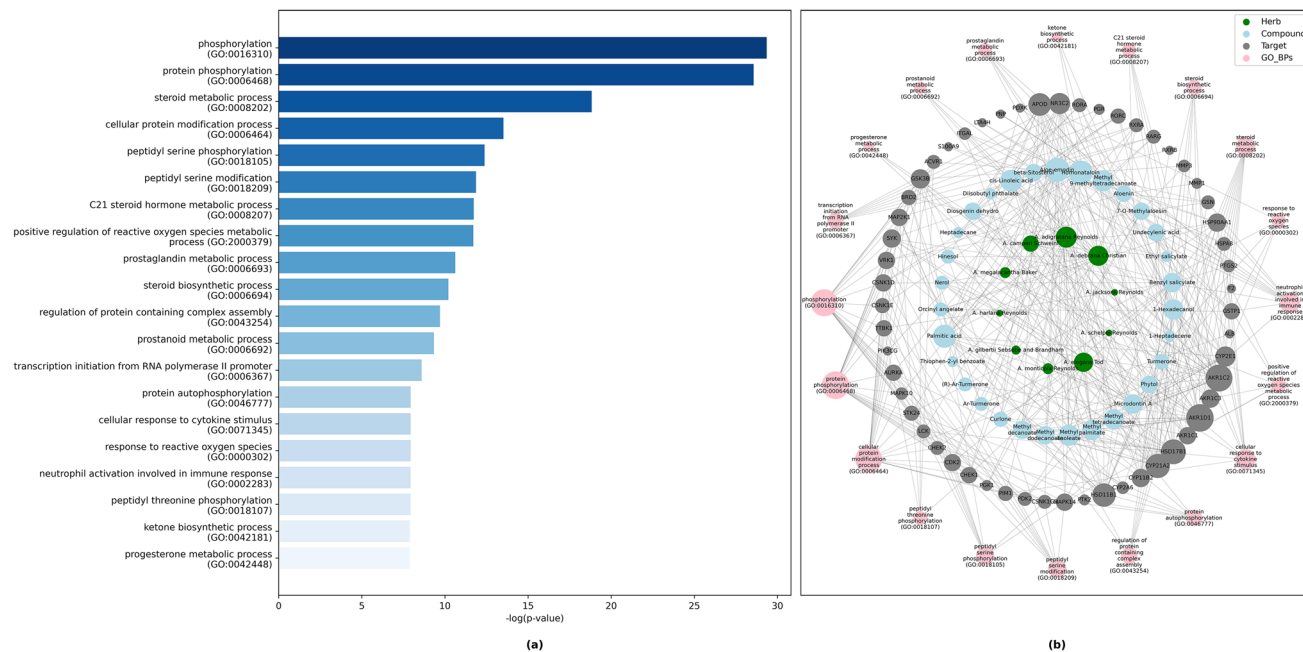
Figure 2. (**a**) Bar graph of KEGG analysis for top 20 enriched biological processes. (**b**) Phytochemical-target-biological process network constructed by python NetworkX library illustrated pathways and their associated genes (Adjusted $P \leq 0.05$).
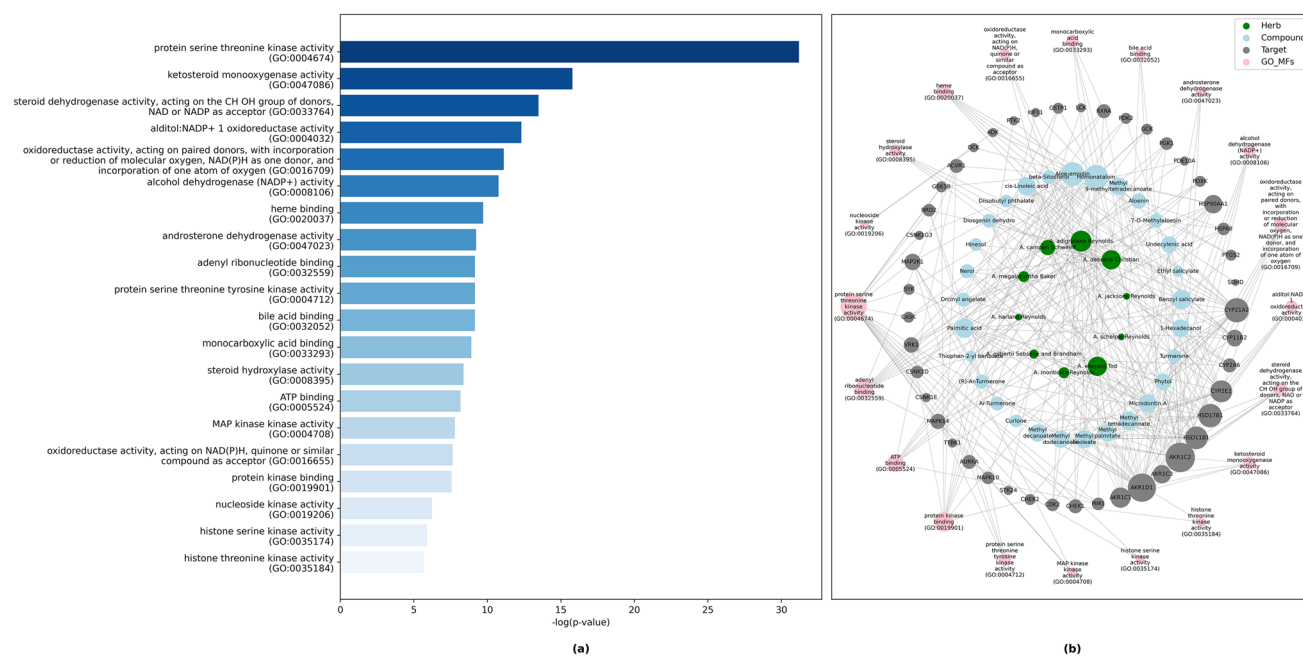

Figure 3. (**a**) Bar graph of KEGG analysis for top 20 enriched molecular functions. (**b**) Phytochemical-target-molecular function network constructed by python NetworkX library illustrated pathways and their associated genes (Adjusted $P \leq 0.05$).

senescence, p53 signaling pathway, small cell lung cancer, and IL-17 signaling pathway and so on. The main enriched BPs for this aloes' phytochemicals targets included: steroid metabolic process, steroid biosynthetic process, protein phosphorylation, transcription initiation from RNA polymerase II promoter, phosphorylation, DNA-templated transcription initiation, glucocorticoid biosynthetic process, glucocorticoid metabolic process, and regulation of cellular carbohydrate metabolic process. Furthermore, the main MFs associated with this aloes' phytochemicals were: protein serine/threonine kinase activity, steroid dehydrogenase activity, heme binding, ketosteroid monooxygenase activity, oxidoreductase activity, steroid hydroxylase activity, and oxysterol binding. In a similar fashion, phytochemicals from other aloe species have related KEGG pathways, BPs and CCs as shown in Supplementary Figure S1 online.
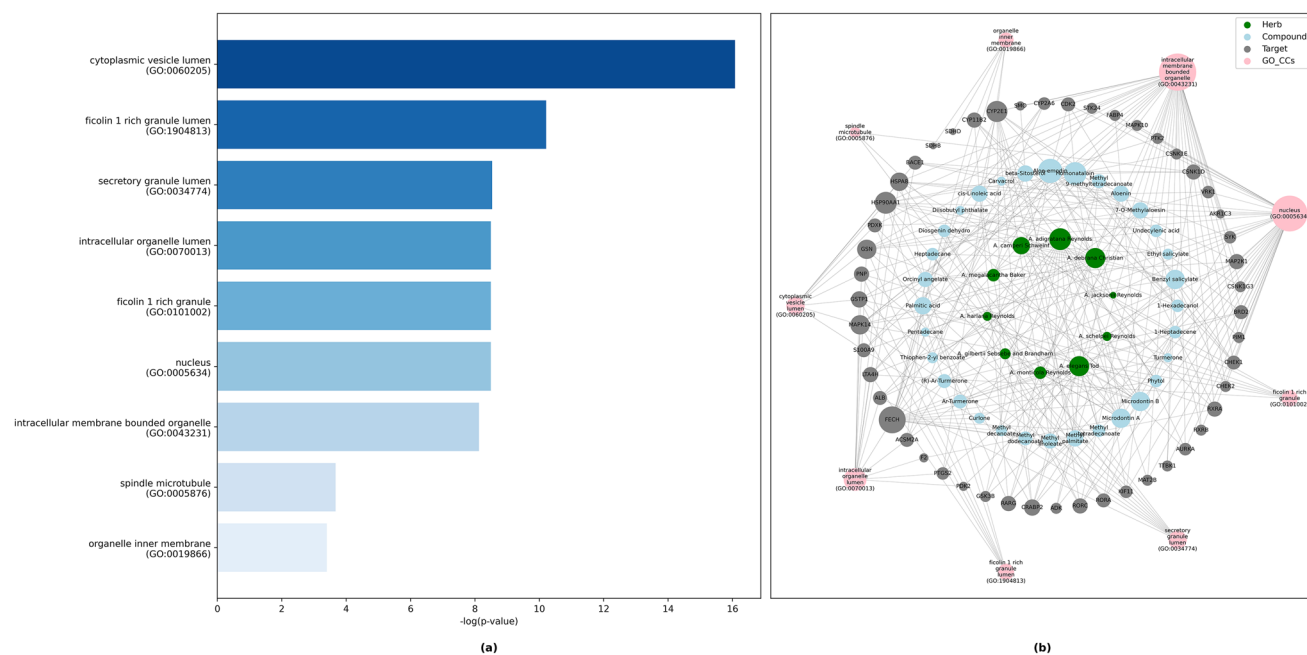
Figure 4. (**a**) Bar graph of KEGG analysis for top 20 enriched cellular components. (**b**) Phytochemical-target-cellular components network constructed by python NetworkX library illustrated pathways and their associated genes (Adjusted $P \leq 0.05$).

By comparing the top 20 biological processes and pathways enrichment data, 42 targets were found in common and the top ten pathways related to these targets were steroid hormone biosynthesis, lipid and atherosclerosis, chemical carcinogenesis, IL-17 signaling pathway, pathways in cancer, Th17 cell differentiation, progesterone-mediated oocyte maturation, prostate cancer, kaposi sarcoma-associated herpesvirus infection, and human T-cell leukemia virus 1 infection.

Potential therapeutic activities of Ethiopian indigenous aloes.

According to pharmacological profiling, phytochemicals from Ethiopian indigenous aloes were linked to cancers (66 targets), endocrine diseases (16 targets), metabolic diseases (15 targets), immune system diseases (5 targets), nervous system diseases (11 targets), congenital disorders of metabolism (3 targets), musculoskeletal diseases (3 targets), digestive system diseases (3 targets), hematologic diseases (2 targets), infectious diseases (3 targets), mouth and dental diseases (1 targets). Some of these pharmacological activities are related to traditionally reported medicinal uses of the aloes.

From our analysis, we found that around 10 cancer types namely adenocarcinoma, breast carcinoma, glioma, malignant neoplasm of breast, malignant neoplasm of prostate, mammary neoplasms, prostatic neoplasms, squamous cell carcinoma of esophagus, leukemia, and melanoma were related to 65 targets that were linked to 34 compounds from all of the 10 aloe species used in this study. Supplementary Table S3 online provides information on the predicted targets and diseases for each phytochemical found in indigenous Ethiopian aloes.

A. adigratana Reynolds phytochemicals were predicted to have therapeutic effect against several cancers (adenocarcinoma, breast carcinoma, malignant neoplasm of prostate, and prostatic neoplasms), endocrine and metabolic diseases (diabetes mellitus, non-insulin-dependent obesity), immune system diseases (colitis, inflammatory bowel diseases and ulcer), digestive system disease (ulcer), and infectious diseases (ulcer and colitis). In northern Ethiopia, *A. adigratana* Reynolds was traditionally used to treat a variety of diseases, including wounds, infections, and inflammation[50]. Previous in vivo and in vitro studies also support the use of *A. adigratana* Reynolds in traditional medicine by confirming its anti-inflammatory properties, which were at least partially related to the presence of aloin A/B and microdontin A/B in the leaves[50].

A. camperi Schweinf. phytochemicals were predicted to have therapeutic effect on cancers (adenocarcinoma, breast carcinoma, malignant neoplasm of breast, malignant neoplasm of prostate, mammary neoplasms, prostatic neoplasms), and endocrine and metabolic diseases (diabetes mellitus- non-insulin-dependent, insulin resistance, morbid obesity, obesity and morbid obesity). Although there is little information on the traditional utilization of the plant, it is claimed to be used for wound healing and malaria treatment[51].

The potential therapeutic activity of the active phytochemicals inferred in this study should be further investigated using in vivo and in vitro scientific experiments on the predicted diseases.

Protein–protein interaction (PPI) network, molecular docking and molecular dynamics (MD) simulations.

There were a total 17,358 protein nodes in the CODA PPI network (Fig. 5a) and the distribution of both degree centrality and eigenvector centrality were examined. The degree centrality values of the top 1, 5 and 10% nodes were greater than or equal to 243, 101 and 67 respectively. The eigenvector centrality
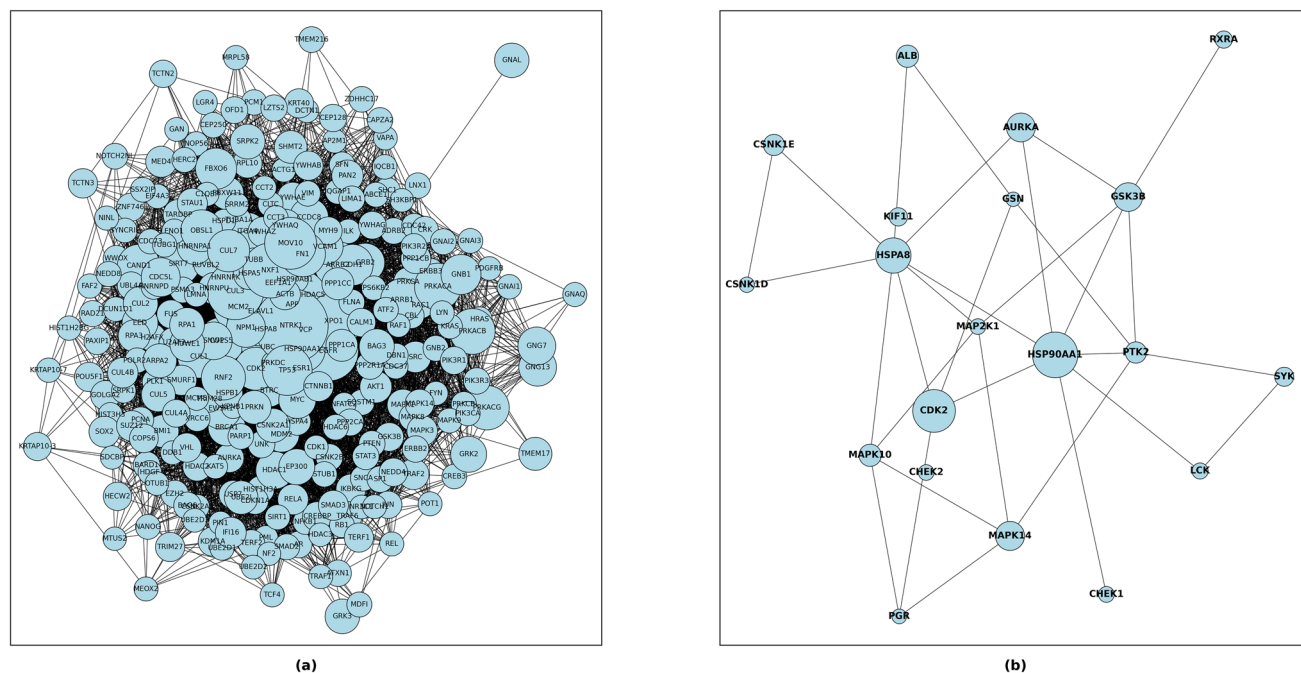
Figure 5. (**a**) PPI network constructed with CODA. (**b**) Network of hub predicted target genes.
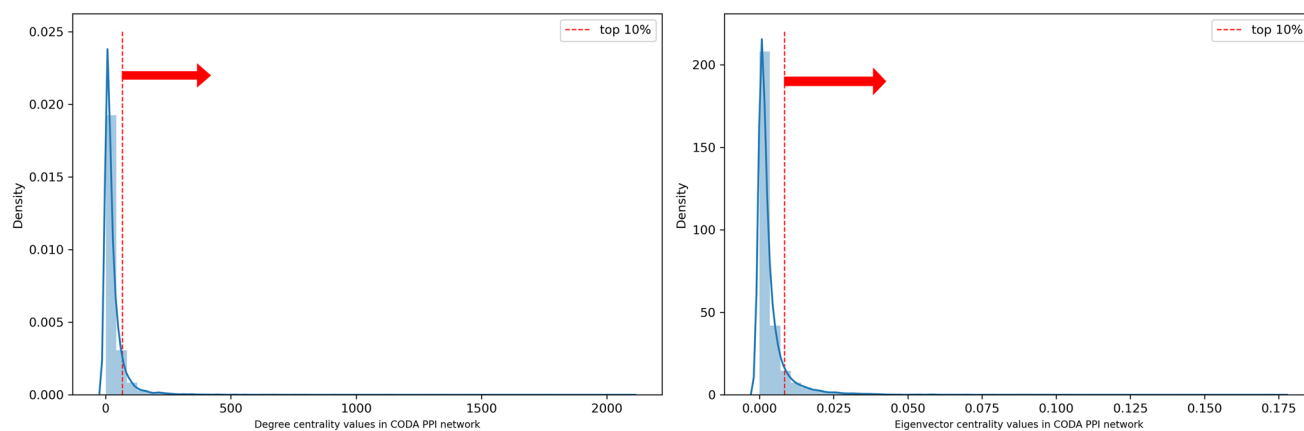

Figure 6. Distribution of centralities of all nodes in protein–protein interaction network. (Left) shows distribution of degree centrality and, (Right) shows distribution of eigenvector centrality.

values were greater than or equal to 0.0313, 0.0138 and 0.008 for 1, 5 and 10% of the nodes respectively. 20 predicted targets (CDK2, HSPAA1, HSPA8, AURKA, MAPK14, GSK3B, MAPK10, PTK2, CSNK1E, SYK, CHEK1, MAP2K1, KIF11, ALB, RXRA, GSN, CHEK2, PGR, CSNK1D, and LSK) were identified to be in the top 10% degree and eigenvector centralities (Fig. 5b). Among them, the targets CDK2, HSPAA1, HSPA8, AURKA, and MAPK14 lies in the top 1% of degree and eigenvector centralities (Fig. 6).

The docking results showed successful binding of the ligands to the target proteins with significantly lower ($p \le 0.05$) binding energies compared to the docking scores of random compounds from PubChem to random targets from PDB, confirming the interactions predicted by pharmacophore models in this study. The binding affinity energies, interacting amino acid residues, docking center and docking size between selected phytochemicals and targets are shown in Table 4. Figure 7 shows the 3D and 2D pose views of the interactions between randomly selected phytochemicals from Ethiopian indigenous aloes and hub gene/protein in PPI network of the predicted targets.

According to docking analysis, the ligand homonataloin was selected for MD simulations as it has relatively minimum binding energy score (−8.7 kcal/mol) with target protein CDK2. MD trajectories over 100 ns for the protein—ligand complex was analyzed to investigate the stability of the proteins with bound ligand using root-mean-square deviation (RMSD) parameter. In order to examine the conformational variation of target protein upon interaction with ligand, RMSD for protein—ligand complex, ligand and target protein with respect to the starting structure was calculated and depicted as shown in Fig. 8. It is clearly shown that RMSD of the atoms of protein – ligand complex, protein and ligand have the same pattern of fluctuation with a steady increase in

Phytochemicals	Target (PDB ID)	Docking score (kcal/mol)	Interacting amino acid residues	Docking center (x, y, z)	Docking size (x, y, z)
Homonataloin	CDK2 (2FVD)	−8.7	ILE10, GLU12, GLY13, VAL18, ALA31, LYS33, VAL64, PHE80, GLU81, PHE82, LEU83, HIS84, GLN85, ASP86, GLN131, ASN132, LEU134, ALA144, ASP145, GLU162	−4, 30, 10	21, 21, 21
Aloenin	HSPA8 (1ATR)	−7.8	TYR41, LYS56, VAL59, PRO63, PHE68, ASP69, HIS89, PRO91, GLU231, ASN235, ARG261, ARG262, ARG264, THR265, GLU268	14, 32, 19	30, 35, 30
Aloe-emodin	AURKA (5DT0)	−8.3	LEU139, GLY140, VAL147, ALA160, LYS162, LEU194, LEU210, GLU211, TYR212, ALA213, GLY216, THR217, GLU260, LEU263, ALA273, ASP274	−6, −37, 4	20, 28, 20
Benzyl-salicylate	MAPK14 (7BDO)	−7.8	VAL30, TYR35, VAL38, ALA51, LYS53, GLU71, LEU75, ILE84, LEU104, THR106, HIS107, LEU108, MET109, LEU167, ASP168, PHE169	18, 32, 18	20, 30, 20
Aloe-emodin	GSK3B (4J71)	−7.9	ILE62, GLY63, VAL70, ALA83, LYS85, GLU97, VAL110, LEU132, ASP133, TYR134, VAL135, PRO136, LEU188, CYS199, ASP200, PHE201	−40, 7, 43	20, 20, 20

Table 4. Binding affinity energies, interacting amino acid residues, docking center and docking size between selected phytochemicals and targets.
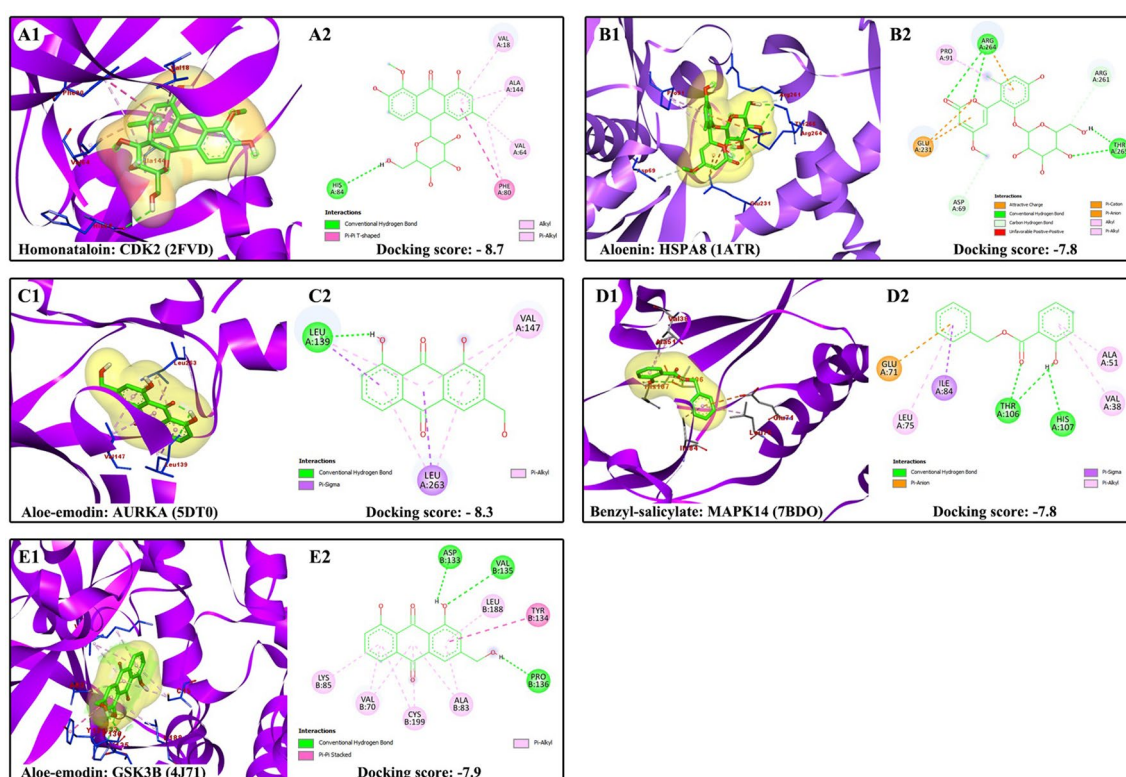
Figure 7. Molecular docking analysis of selected phytochemicals and targets. (**A**1-**E**1) 3D pose views of interaction selected phytochemicals and targets. (**A**2-**E**2) 2D pose views of interaction of selected phytochemicals and targets.

the initial 0.25 ns followed by stability until the end of the simulation. The RMSD of the protein (Fig. 8, red) is slightly higher than that of the ligand (Fig. 8, black) during the course of simulation. Furthermore, the RMSD of the protein—ligand complex (Fig. 8, green) is higher than that of the protein and ligand indicating that there was a conformational change of protein in the complex with the ligand binding. Interestingly, the protein—ligand complex exhibited a very low fluctuations during the simulation period indicating that the complex is stable with the ligand remaining fit in the binding pocket of the receptor (see Supplementary Video S1 online).

An excellent analysis of the dynamics of protein collapse is provided by the time development of the radius of gyration (Rg)[52]. To determine whether the protein was compact, the systems' gyration radii were measured, and a graph showing how they are related to the simulation time was created[53] as shown in Fig. 9. Similar to RMSD plots, Rg of the ligand, protein and the protein – ligand complex system did not change significantly throughout the simulation indicating that the binding region shows little influence on their structures during the simulation period.
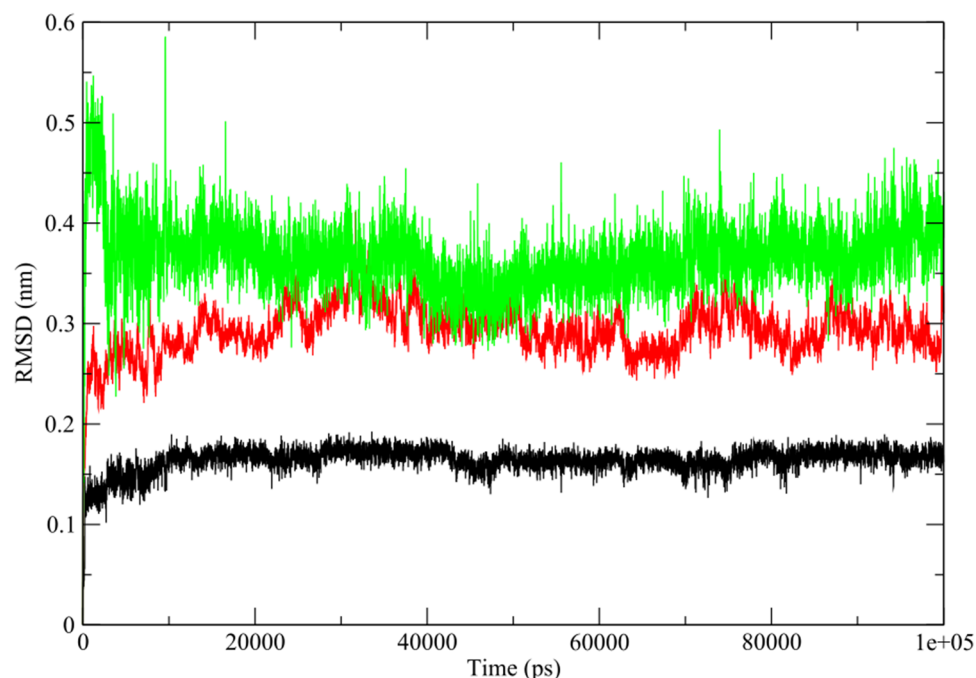
Figure 8. RMSD plots of target proteins (CDK2)—ligand (homonataloin) complex over 100 ns MD simulation.
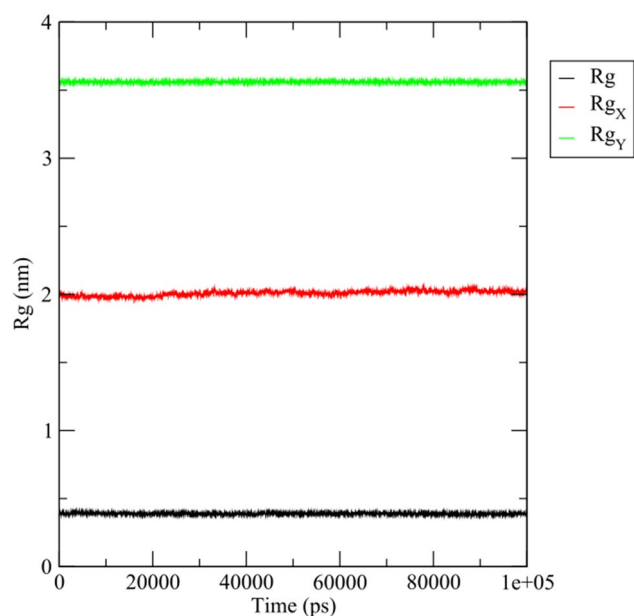

Figure 9. Radius of gyration, Rg (nm) plots of the ligand (Rg), target protein (Rg_X) and protein–ligand complex (Rg_Y) over 100 ns simulation run.

Methods

In this study, pharmacophore model-based ligand profiling was used in target fishing to predict potential targets of compounds from Ethiopian indigenous aloes. Figure 10 shows the workflow of this study. First, Ethiopian indigenous aloes were identified through literature searches and their phytochemicals (compounds) were manually compiled from literatures and integrated Ethiopian traditional herbal medicine and phytochemicals database (ETM-DB)[54]. The three-dimensional (3D) structure data files (SDF) of the phytochemicals were collected from PubChem database[55]. The physicochemical properties of the phytochemicals were retrieved from PubChem database and SwissADME webserver[29]. The phytochemicals were filtered for toxicity using admetSAR. Then, we used PharmaDB pharmacophore and target databases of Discovery Studio[56] to perform target fishing for the phytochemicals. The hit pharmacophore models were picked out according to the threshold of a predetermined fit
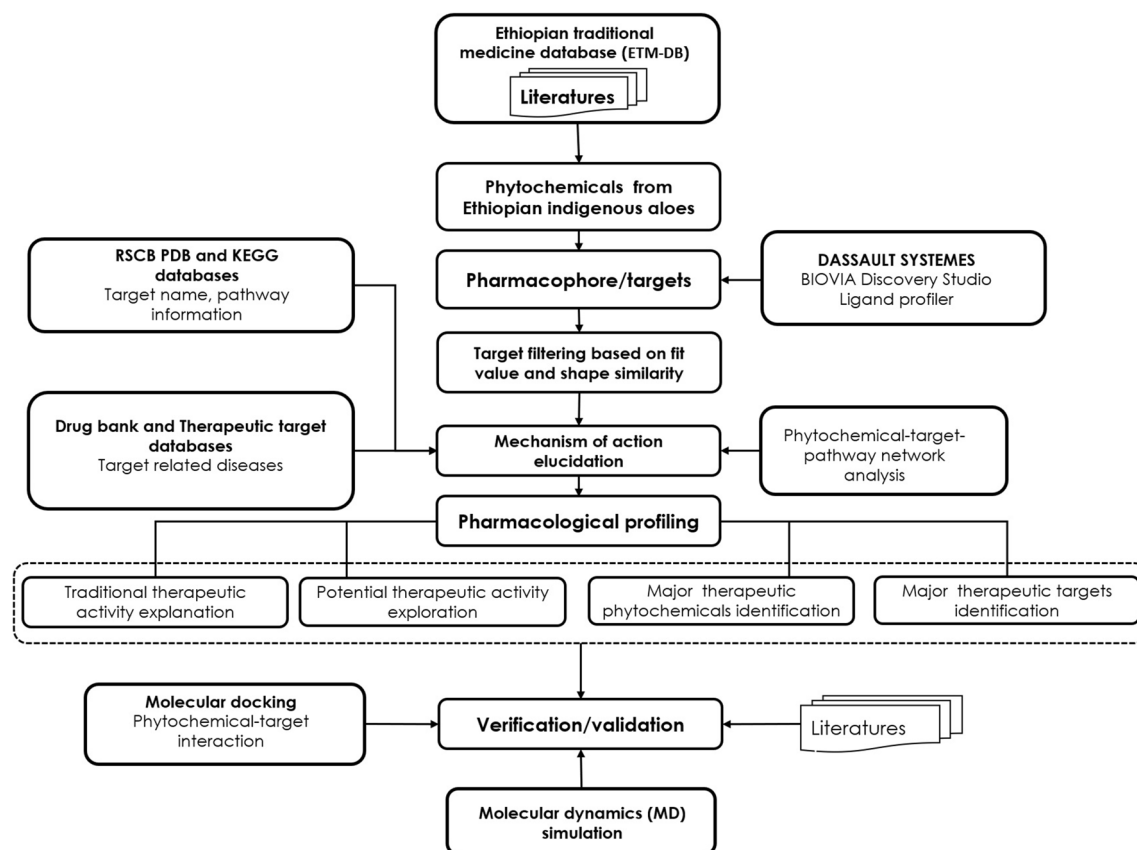
Figure 10. The workflow used in this study.

value (0.5). After analysis of the hit targets and their associated pathways and diseases, as well as the interactions between the phytochemicals and the targets, an action network of Ethiopian indigenous aloes phytochemicals was constructed. Furthermore, Molecular docking was carried out to refine the binding modes and to verify the findings. Finally, molecular dynamics (MD) simulation was performed for the best binding interaction of the protein—ligand complex.

Collection of phytochemical constituents of Ethiopian indigenous aloes. The phytochemical constituents of Ethiopian indigenous aloes were collected from our own database[54] and the literature. Accordingly, we compiled 35 compounds from 10 Ethiopian indigenous aloes after removing potentially toxic compounds and evaluating their ADMET and drug-likeness properties. And then, targets were successfully predicted for the phytochemicals using computational target fishing. All 35 phytochemicals' 2D chemical structures and compound classes are listed in Fig. 11. These 35 phytochemicals were classified using ClassyFire (automated chemical classification with a comprehensive, computable taxonomy)[57] into 9 compound classes as anthracenes, benzene and substituted derivatives, fatty acyls, organooxygen compounds, phenol esters, prenol lipids, saturated hydrocarbons, steroids and steroid derivatives, and unsaturated hydrocarbons.

Evaluation of physicochemical, drug-likeness and ADMET properties of phytochemicals from Ethiopian indigenous aloes. Physicochemical properties of the phytochemicals were retrieved from PubChem database[28] and SwissADME webserver[29]. SwissADME is a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. We used PubChemPy—a python library used to interact with PubChem database to retrieve information related to phytochemicals from PubChem database.

The drug-likeness properties of the phytochemicals were predicted using the physicochemical properties. Drug-likeness is related to molecular properties affecting pharmacodynamics and pharmacokinetics of the molecules[58]. These molecular properties are related to some basic structural or physicochemical properties such as $\log p$ (partition coefficient), molecular weight (MW), topological polar surface area (TPSA), hydrogen bond acceptors and donors count in a molecule. There are several sets of criteria to select ligands for a potential drug. Most notably, Lipinski[30] identified common properties that are frequently observed in approved compounds. The authors presented a "rule of 5" for drug-like molecules (less than 5 hydrogen bond donors, less than 10 hydrogen bond acceptors, less than 500 Daltons (g/mol) molecular weight, and less than 5 $\log p$ (partition coefficient between organic and aqueous phases)). In addition, Veber's rule[31] suggested that compounds which meet only the two criteria of 10 or fewer rotatable bonds and polar surface area equal to or less than 140 Å^2 have a
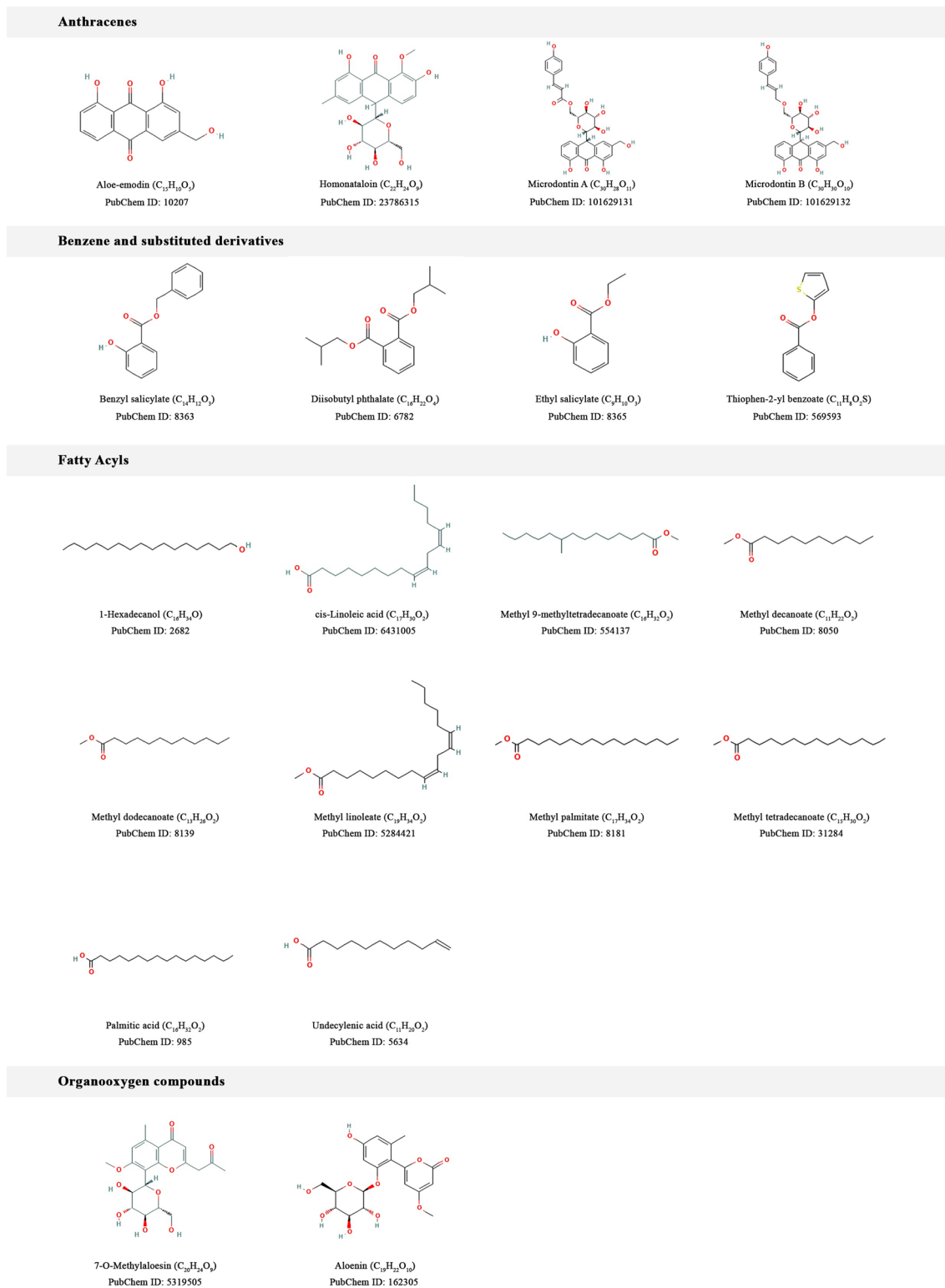
Figure 11. 2D chemical structures and compound classes of phytochemical constituents of Ethiopian indigenous aloes whose human targets were successfully predicted using pharmacophore models.

high probability of good oral bioavailability in rat. However, many of the most successful drugs do not fit these guidelines, and care should be taken in the application of these guidelines[32]. In addition, most of compounds from natural products do not always follow the rules which are commonly applied in modern drug design including Lipinski's "Rule of Five "[21].We used compounds with two or more Rule of 5 violations which are used in drug development by medicinal chemists and industries[59] as a reference for drug-likeness evaluation of the phytochemicals in this research.

Phenol esters
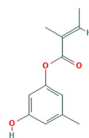
Orcinyl angelate (C$_{12}$H$_{14}$O$_2$)
PubChem ID: 91692591

Prenol lipids
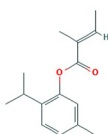
(E)-2-Isopropyl-5-methylphenyl 2-methylbut-2-enoate
(C$_{15}$H$_{20}$O$_2$)
PubChem ID: 91698167
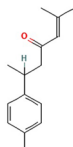
(R)-Ar-Turmerone (C$_{15}$H$_{20}$O)
PubChem ID: 558221
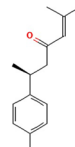
Ar-Turmerone (C$_{15}$H$_{20}$O)
PubChem ID: 160512
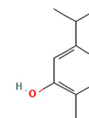
Carvacrol (C$_{10}$H$_{14}$O)
PubChem ID: 10364
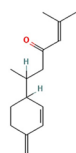
Curlone (C$_{15}$H$_{22}$O)
PubChem ID: 196216
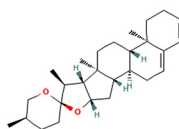
Diosgenin dehydro (C$_{27}$H$_{40}$O$_2$)
PubChem ID: 337494
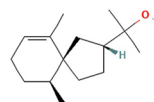
Hinesol (C$_{15}$H$_{26}$O)
PubChem ID: 10878761
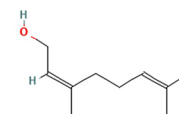
Nerol (C$_{10}$H$_{18}$O)
PubChem ID: 643820
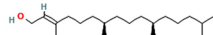
Phytol (C$_{20}$H$_{40}$O)
PubChem ID: 5280435
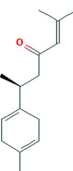
Turmerone (C$_{15}$H$_{22}$O)
PubChem ID: 14367555

Saturated hydrocarbons
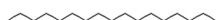
Heptadecane (C$_{17}$H$_{36}$)
PubChem ID: 12398

Pentadecane (C$_{15}$H$_{32}$)
PubChem ID: 12391

Steroids and steroid derivatives
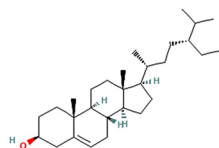
Unsaturated hydrocarbons

Beta-Sitosterol (C$_{29}$H$_{50}$O)
PubChem ID: 222284

1-Heptadecene (C$_{17}$H$_{34}$)
PubChem ID: 23217

Figure 11. (continued)

ADMET properties were predicted from admetSAR web server[27] from the compounds' Simplified Molecular Input Line Entry System (SMILES) string as input using BeautifulSoup (a Python library for pulling data out of HTML and XML file). We removed potentially toxic compounds that were AMES mutagens, carcinogens and human Ether-a-go-go-Related gene (hERG) inhibitors. The inhibition of the human ether-a-go-go (hERG) ion channel may cause cardiotoxicity and hence, it is necessary to screen all potential drug candidates for risk against blocking the hERG channel[60]. The Ames test is a sensitive tool in screening for potential genotoxic carcinogens[61].

Pharmacophore model-based in silico profiling of phytochemicals from Ethiopian indigenous aloes. A pharmacophore model represents a target-ligand binding site that triggers a desired pharmacological effect[16]. A pharmacophore models used in this study were derived from protein–ligand 3D complex structures as well as structural data on small bioactive organic molecules[62]. Pharmacophores can be used as low-dimensional targets in target prediction tasks[63]. We performed pharmacophore-based prediction of the potential targets of aloes' phytochemicals using BIOVIA Discovery Studio[56]. Discovery Studio is an interactive computational chemistry and molecular modeling package with a broad variety of features for modeling and simulation[56]. We used "Ligand Profiler" protocol which is an automated ligand profiling in BIOVIA Discovery Studio. The BIOVIA Discovery Studio provides automated pharmacophore-based activity profiling and reporting[16]. Discovery Studio 4.5 is equipped with PharmaDB pharmacophore database which contains about 68,000 pharmacophores derived from 8000 protein–ligand complexes from the sc-PDB dataset, which itself is a collection of 3D structures of binding site found in PDB[64]. The binding sites were extracted from crystal structures in which a complex between a protein cavity and a small molecule ligand could be identified[15].

As for parameters setting during ligand profiling with Input PharmaDB Pharmacophores, all the pharmacophore models with the shape of the binding pocket were selected for the virtual screening with default settings. The RIGID mode was used as the molecular mapping algorithm. A conformer database (3D SDF file format) was generated using BEST settings of the Conformation Generator component for each query ligand. All molecular features were allowed while mapping these ligands to the pharmacophore models to increase selectivity. The minimal interfeature distance was set at 0.5 Å. For every query ligand, targets were ranked by decreasing fit-value. The targets were nominated according to their UniProt ID. UniProt is an abbreviation for the Universal Protein Database program, which consists of the Swiss-Prot, TrEMBL and PIR-PSD databases and is the largest database containing the most informative data resources and protein structures[65]. Further information about the protein including the protein name, gene name, organism, sequence information, taxonomy, family and domains, and related molecular functions and biological processes can be obtained using the UniProt ID from UniProt website (http://www.uniprot.org/). This database can also link to other databases by searching based on the PDB ID, KEGG ID and other information. In this work, for each target, the target name, UniProt entry name and UniProt ID information were collected from the PDB[64] and UniProt databases[65].

Inferring the pharmacological actions using KEGG, GO and network analysis. To investigate the therapeutic effect of Ethiopian indigenous aloes, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene ontology (GO) of the 82 predicted targets were analyzed. In order to identify the complex relationship between compounds, targets, pathways and biological processes, we conducted enrichment test by using enrichr function in gseapy python library with internal gene set named 'KEGG_2021_Human'. Network analysis in the Context-Oriented Directed Associations (CODA)[42] were performed to obtain diseases related to each predicted targets. Network propagation was conducted with Random-Walk with Restart (RWR) algorithm in the heterogeneous biological networks of CODA. CODA is an integrated database which contains diverse types of biological interactions from molecular level to phenotypic level. We constructed a heterogeneous network using all types of biological interactions in CODA to see converging information flow into disease nodes in the network starting from the targets. After running the algorithm, we ranked disease nodes in the network by their RWR scores. Then the disease was mapped to high-level disease category using KEGG Application Programming Interface (API)—a REST-style API to the KEGG database resource.

Protein–protein interaction (PPI) network, molecular docking and molecular dynamics (MD) simulations. A protein–protein interaction (PPI) network was constructed from CODA[42], which contains a total of 17,358 protein nodes. Among molecular relations in the CODA, proteins with both left and right entity type of gene product were selected. Then all the nodes were represented with gene symbol for better interpretation. To get hub targets in the PPI network, two types of network centralities (degree centrality and eigenvector centrality) were used. There are also another two more representative network centrality measures such as closeness centrality and betweenness centrality. However, closeness and betweenness centrality were not considered in this study because they focus on the overall robustness in the network than relationship between neighboring nodes[66]. Because of this, we consider hub targets as the nodes within the top 10% degree and eigenvector centrality values in the PPI network.

Additionally, molecular docking studies were performed to confirm the protein–ligand interaction that the pharmacophore models predicted. Blind docking was used for the detection of possible binding sites and modes of peptide ligands by scanning the entire surface of protein targets[67]. In this study, we employed blind docking since we do not have information regarding the binding mode of the phytochemicals with the predicted targets. The PDB files of five randomly chosen proteins from hub proteins in PPI network (obtained from CODA network analysis) were retrieved from Protein Data Bank (PDB), and the ligands' 3D SDF file was obtained from PubChem database. The molecular docking was done using the CB-Dock2 web server (http://clab.labshare.cn/cb-dock/php/index.php), which is a server for protein–ligand blind docking, integrating cavity detection, docking and homologous template matching. CB-Dock automatically identifies the binding sites, calculates the center and size, customizes the docking box size according to the query ligand and then perform the molecular docking with AutoDock Vina[68]. Then, the 3D and 2D poses of the protein—ligand complexes were analyzed and visualized using Discovery Studio Visualizer.

Furthermore, we compared the docking score of our selected target and protein with the docking scores of random compounds from PubChem docked with random targets from PDB using one-sample, one-sided t-test to statistically show that the bindings were significant ($P \leq 0.05$).

Among the selected phytochemicals for molecular docking, the ligand homonataloin displayed the best binding interaction and free energy with the Cyclin Dependent Kinase 2 (CDK2) target protein. Accordingly, it was chosen to be subjected to 100 nanoseconds (ns) molecular dynamics (MD) simulation against the target protein to confirm its stable binding affinity. MD simulations were performed using GROningen MAchine for Chemical Simulations (GROMACS) 2021.4 software package[69] where the CHARMM27 all-atom forcefield[70] was used for protein topology preparation and the official SwissParam (https://www.swissparam.ch/)[71] for ligand topology preparation. The solvation method used was a triclinic box of default Simple Point Charge (SPC) water model where explicit solvent periodic boundary conditions were applied. Charge neutralization using sodium and chloride ions was carried out for the solvated complexes. The systems were subjected to energy minimization to resolve any steric clashes or inappropriate geometry employing the steepest descent method through 5000 steps. System equilibration was also manipulated to ensure a reasonable starting structure using NVT; equilibration under constant number of particles, volume, and temperature (NVT) for 100 picoseconds (ps) using a Berendsen thermostat[72]. Also, re-equilibration was performed for another 100 ps under constant pressure (Isothermal-isobaric (NPT) ensemble) using the Parrinello-Rahman barostat using a time step of 2 femtoseconds (fs) for each equilibration round[73]. Finally, MD production phase was carried out for 100 ns using a time step of 2 fs at a constant temperature of 300 K and constant pressure of 1 atm. For MD trajectories analysis, the complex was re-centered and re-wrapped within the unit cells using the "trjconv" function of GROMACS. Then, the trajectories were analyzed using RMSD of the protein, ligand and protein–ligand complex referenced to its initial position and the radius of gyration (Rg). GRaphing and Advanced Computation and Exploration of data (Grace) tool was used for data plotting (https://plasma-gate.weizmann.ac.il/Grace/).

Computation environment. The BIOVIA Discovery Studio software program were installed on Windows 10. Profiling of the phytochemicals were carried out in an automatic fashion on a Windows 10 computer (Intel Core i5-4690 CPU @ 3.50 GHz, 3501 MHz, 4 Core(s), and 4 Logical Processor(s)). The configuration of the RAM was: 4 slots, 3 of them each having 8 GB RAM and 1 slot 4 GB RAM giving a combined 28 GB, DDR3 and speed of 1600 MHz. Network analysis (Random-Walk with Restart algorithm), KEGG pathway enrichment and GO analysis were performed on a server CentOS computer, Intel Xeon CPU E5-2680 v3 @ 2.50 GHz, 48 cores. The configuration of the RAM was: 16 slots with 32 GB RAM, DDR4. MD simulation was done using: Linux system (Ubuntu 2021.4–2) computer (Intel Core i7-10,700 CPU @ 2.90 GHz × 16, Mesa Intel UHD Graphics 630 (CML GT2) / llvmpipe (LLVM 13.0.1, 256 bits, 64.0 GB RAM)).

Conclusions

In silico profiling provides a rapid approach for identification of active compounds, prediction of bioactivities of compounds, and to guide the rationalization of the use of traditional medicines. Computational target fishing is a new approach in pharmaceutical research that helps to examine the relationship between drugs and diseases. Using the admetSAR and SwissADME online servers, the ADMET and drug-likeness characteristics of phytochemicals from indigenous aloes in Ethiopia were assessed. Then, using in silico pharmacophore models, potential therapeutic targets were predicted and a wide range of associated therapeutic potential was revealed.

The findings suggested that the phytochemicals in indigenous Ethiopian aloes are potential therapeutic molecules with 82 associated targets that were enriched in a number of KEGG pathways and GO terms. The results demonstrated that the phytochemicals have a broad range of therapeutic targets and the potential to regulate important biological pathways, such as the metabolism of steroid hormone biosynthesis, lipid and atherosclerosis, chemical carcinogenesis, IL-17 signaling pathway, Th17 cell differentiation, and pathways in cancer, among others. Furthermore, this study revealed that the phytochemicals may be used as a potential therapeutic agent for a number of cancers, and diseases of the endocrine, metabolic, immune, nervous, digestive, musculoskeletal, and hematologic systems, as well as congenital disorder of metabolism, infectious diseases, and mouth and dental diseases. Particularly, the aloe species *A. debrana* Christian and *A. megalacantha* Baker were linked cancer and endocrine disease; *A. gilbertii* Sebsebe and Brandham, *A. monticola* Reynolds, and *A. schelpei* Reynolds were linked with congenital disorders of metabolism; *A. adigratana* Reynolds and *A. elegans* Tod were related to digestive system disease; and so forth. This information indicated that these aloes were predicted to be effective in treating the aforementioned diseases. In order to confirm these effects, the phytochemicals' pharmacological effects should be applied to further controlled experimental research.

Data availability

All relevant data are within the paper and supporting information files.

References

1. Medcalf, A., Bhattacharya, S., Momen, H., Saavedra, M. & Jones, M. Health for all : The journey of universal healt coverage. (Orient Blackswan Pvt. Ltd 2015).
2. Mushtaq, S., Abbasi, B. H., Uzair, B. & Abbasi, R. Natural products as reservoirs of novel therapeutic agents. *EXCLI J.* **17**, 420–451 (2018).
3. Ekor, M. The growing use of herbal medicines: Issues relating to adverse reactions and challenges in monitoring safety. *Front. Neurol.* **4**, 1–10 (2014).
4. Cragg, G. M. & Newman, D. J. Natural product drug discovery in the next millennium. *Pharm. Biol.* **39**, 8–17 (2001).
5. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

6. Rainsford, K. D., Powanda, M. C. & Whitehouse, M. W. *Novel natural products: Therapeutic effects in pain, arthritis and gastro-intestinal diseases. Progress in drug research* Vol. 70 (Springer Basel, 2015).

7. Grace, O. M., Simmonds, M. S. J., Smith, G. F. & Van Wyk, A. E. Documented utility and biocultural value of *Aloe* L. (Asphode-laceae): A review. *Econ. Bot.* **63**, 167–178 (2009).

8. Dagne, E., Bisrat, D., Viljoen, A. & Van Wyk, B.-E. Chemistry of Aloe Species. *Curr. Org. Chem.* **4**, 1055–1078 (2005).

9. Belayneh, A., Demissew, S., Bussa, N. F. & Bisrat, D. Ethno-medicinal and bio-cultural importance of aloes from south and east of the Great Rift Valley floristic regions of Ethiopia. *Heliyon* **6**, e04344 (2020).

10. Oldfield, S. A. Review of significant trade: East African aloes. Document 9.2. 2 Annex 4. In: *14th meeting of the CITES Plants Committee, Windhoek Namibia* 16–20 (2004).

11. Demissew, S., Nordal, I. & Stabbetorp, O. E. Endemism and patterns of distribution of the genus Aloe (Aloaceae) in the flora of Ethiopia and Eritrea. *Biol. Skr.* (2001).

12. Edwards, S., Nemomissa, S. & Hedberg, I. *Flora of Ethiopia and eritrea* (Addis Ababa University, 2003).

13. Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M. & Taranto, A. G. Structure-based virtual screening: from classical to artificial intelligence. *Front. Chem.* **8**, 343 (2020).

14. Hillisch, A., Pineda, L. F. & Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug Discov. Today* **9**, 659–669 (2004).

15. Meslamani, J. *et al.* Protein-ligand-based pharmacophores: Generation and utility assessment in computational ligand profiling. *J. Chem. Inf. Model.* **52**, 943–955 (2012).

16. Schuster, D. 3D pharmacophores as tools for activity profiling. *Drug Discov. Today Technol.* **7**, e205–e211 (2010).

17. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J. & Bork, P. Drug target identification using side-effect similarity. *Science* **321**, 263–266 (2008).

18. Dubus, E., Ijjaali, I., Barberan, O. & Petitet, F. Drug repositioning using in silico compound profiling. *Future Med. Chem.* **1**, 1723–1736 (2009).

19. Anighoro, A., Bajorath, J. & Rastelli, G. Polypharmacology: Challenges and opportunities in drug discovery. *J. Med. Chem.* **57**, 7874–7887 (2014).

20. Wermuth, C. G., Ganellin, C. R., Lindberg, P. & Mitscher, L. A. Glossary for chemists of terms used in medicinal chemistry. *Pure Appl. Chem.* **70**, 1129–1143 (1998).

21. Seidel, T., Wieder, O., Garon, A. & Langer, T. Applications of the pharmacophore concept in natural product inspired drug Design. *Mol. Inform.* **39**, e2000059–e2000059 (2020).

22. Patel, Y., Gillet, V. J., Bravi, G. & Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput. Aided. Mol. Des.* **16**, 653–681 (2002).

23. Qing, X. *et al.* Pharmacophore modeling: Advances, limitations, and current utility in drug discovery. *J. Receptor. Ligand Channel Res.* **7**, 81–92 (2014).

24. Guner, O. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.* **2**, 1321–1332 (2005).

25. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

26. Lucas, A. J., Sproston, J. L., Barton, P. & Riley, R. J. Estimating human ADME properties, pharmacokinetic parameters and likely clinical dose in drug discovery. *Expert Opin. Drug Discov.* **14**, 1313–1327 (2019).

27. Cheng, F. *et al.* AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* **52**, 3099–3105 (2012).

28. Kim, S. *et al.* PubChem substance and compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2015).

29. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 42717 (2017).

30. 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.* **64**, 4–17 (2012).

31. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).

32. Cosconati, S. *et al.* Virtual screening with autodock: Theory and practice. *Expert Opin. Drug Discov.* **5**, 597–607 (2010).

33. Shannalee R. Martinez, Maresha S. Gay, and L. Z. BDDCS, the Rule of 5 and Drugability. *Adv. Drug Deliv. Rev.* **176**, 139–148 (2016).

34. Palm, K., Stenberg, P., Luthman, K. & Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **14**, 568–571 (1997).

35. Jagannathan, R. Characterization of drug-like chemical space for cytotoxic marine metabolites using multivariate methods. *ACS Omega* **4**, 5402–5411 (2019).

36. Doak, B. C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. *Chem. Biol.* **21**, 1115–1142 (2014).

37. Azman, M., Sabri, A. H., Anjani, Q. K., Mustaffa, M. F. & Hamid, K. A. Intestinal absorption study: Challenges and absorption enhancement strategies in improving oral drug delivery. *Pharmaceuticals* **15**, 1–24 (2022).

38. Shirasaka, Y., Sakane, T. & Yamashita, S. Effect of P-glycoprotein expression levels on the concentration-dependent permeability of drugs to the cell membrane. *J. Pharm. Sci.* **97**, 553–565 (2008).

39. Amin, M. L. P-glycoprotein Inhibition for optimal drug delivery. *Drug Target Insights* **7**, 27–34 (2013).

40. Aller, S. G. *et al.* Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **323**, 1718–1722 (2009).

41. Lynch, T. & Price, A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am. Fam. Physician* **76**, 391–396 (2007).

42. Yu, H. *et al.* CODA: Integrating multi-level context-oriented directed associations for analysis of drug effects. *Sci. Rep.* **7**, 7519 (2017).

43. Schiffer, L. *et al.* Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. *J. Steroid Biochem. Mol. Biol.* **194**, 105439 (2019).

44. Davis Robert, H. Method of using aloe vera as a biological vehicle. *United States patent, US 5,708,038* (1998).

45. Dana, N., Javanmard, S. H., Asgary, S., Asnaashari, H. & Abdian, N. The effect of Aloe vera leaf gel on fatty streak formation in hypercholesterolemic rabbits.. *J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci.* **17**, 439–442 (2012).

46. Sanders, B. *et al.* Anti-cancer effects of aloe-emodin: A systematic review. *J. Clin. Transl. Res.* **3**, 283–296 (2018).

47. Manirakiza, A., Irakoze, L. & Manirakiza, S. Aloe and its effects on cancer: A narrative literature review. *East African Heal. Res. J.* **5**, 1–16 (2021).

48. Harris, M. A. *et al.* The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).

49. Consortium,T. G. O. The gene ontology resource: 20 years and still Going strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

50. Tsegay, M., Tewabe, Y., Bisrat, D. & Asres, K. In vivo anti-inflammatory activity of two anthrones from the leaf latexes of Aloe adigratana reynolds and Aloe elegans todaro. *Ethiop. Pharm. J.* **34**, 1–8 (2018).

51. Fentaw, E. *et al.* Uses and perceived sustainability of *Aloe* L. (Asphodelaceae) in the central and northern Highlands of Ethiopia. *South African J. Bot.* **147**, 1042–1050 (2022).

52. Zhan, F. *et al.* Towards understanding the interaction of β-lactoglobulin with capsaicin: Multi-spectroscopic, thermodynamic, molecular docking and molecular dynamics simulation approaches. *Food Hydrocoll.* **105**, 105767 (2020).
53. Sahihi, M. & Ghayeb, Y. An investigation of molecular dynamics simulation and molecular docking: Interaction of citrus flavonoids and bovine β-lactoglobulin in focus. *Comput. Biol. Med.* **51**, 44–50 (2014).
54. Bultum, L. E., Woyessa, A. M. & Lee, D. ETM-DB: Integrated Ethiopian traditional herbal medicine and phytochemicals database. *BMC Complement. Altern. Med.* **19**, 1–11 (2019).
55. Bolton, E. E., Wang, Y., Thiessen, P. A. & Bryant, S. H. PubChem substance and PubChem compound. *Annu. Rep. Comput. Chem.* **4**, 217–241 (2008).
56. Dassault Systèmes. BIOVIA, Discovery Studio. (2021).
57. Steinbeck, C. *et al.* ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 1–20 (2016).
58. Mishra, H., Singh, N., Lahiri, T. & Misra, K. A comparative study on the molecular descriptors for predicting drug-likeness of small molecules. *Bioinform.* **3**, 384–388 (2009).
59. Benet, L. Z., Hosey, C. M., Ursu, O. & Oprea, T. I. BDDCS, the rule of 5 and drugability. *Adv. Drug Deliv. Rev.* **101**, 89–98 (2016).
60. Hanser, T., Steinmetz, F. P., Plante, J., Rippmann, F. & Krier, M. Avoiding hERG-liability in drug design via synergetic combinations of different (Q)SAR methodologies and data sources: A case study in an industrial setting. *J. Cheminform.* **11**, 9 (2019).
61. Nicolette, J. Genetic toxicology testing. in *A Comprehensive Guide to Toxicology in Nonclinical Drug Development.* 129–154 (Elsevier, 2017).
62. Sydow, D. *et al.* Advances and challenges in computational target prediction. *J. Chem. Inf. Model.* https://doi.org/10.1021/acs.jcim.8b00832 (2019).
63. McGregor, M. J. & Muskal, S. M. Pharmacophore fingerprinting. 2. Application to primary library design. *J. Chem. Inf. Comput. Sci.* **40**, 117–125 (2000).
64. Meslamani, J., Rognan, D. & Kellenberger, E. sc-PDB: A database for identifying variations and multiplicity of 'druggable' binding sites in proteins. *Bioinformatics* **27**, 1324–1326 (2011).
65. Consortium U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699 (2018).
66. MD Aksam, V. K., Chandrasekaran, V. M. & Pandurangan, S. Hub nodes in the network of human mitogen-activated protein kinase (MAPK) pathways: Characteristics and potential as drug targets. *Informatics Med. Unlocked* **9**, 173–180 (2017).
67. Hetényi, C. & van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **580**, 1447–1450 (2006).
68. Liu, Y. *et al.* CB-Dock: A web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacol. Sin.* **41**, 138–144 (2020).
69. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
70. Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B. & Lindahl, E. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **6**, 459–466 (2010).
71. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).
72. Golo, V. L. & Shaĭtan, K. V. Dynamic attractor for the Berendsen thermostat an the slow dynamics of biomacromolecules. *Biofizika* **47**, 611–617 (2002).
73. Tuble, S. C., Anwar, J. & Gale, J. D. An approach to developing a force field for molecular simulation of martensitic phase transitions between phases with subtle differences in energy and structure. *J. Am. Chem. Soc.* **126**, 396–405 (2004).

Acknowledgements

Author contributions

L.E.B. conceptualized and designed the research, wrote the manuscript, performed the experiments and analyzed the data; G.B.T. designed the research; G.K. processed and analyzed the data, O.K. performed data processing and visualization, D.L. designed the research and supervised the overall project. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-26446-x.

Correspondence and requests for materials should be addressed to D.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.