MDPI

Article

Clustering Analysis, Structure Fingerprint Analysis, and Quantum Chemical Calculations of Compounds from Essential Oils of Sunflower (*Helianthus annuus* L.) Receptacles

Yi He, Kaifeng Liu, Lu Han * and Weiwei Han *

Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, School of Life Science, Jilin University, 2699 Qianjin Street, Changchun 130012, China
* Correspondence: luhan@jlu.edu.cn (L.H.); weiweihan@jlu.edu.cn (W.H.)

Abstract: Sunflower (*Helianthus annuus* L.) is an appropriate crop for current new patterns of green agriculture, so it is important to change sunflower receptacles from waste to useful resource. However, there is limited knowledge on the functions of compounds from the essential oils of sunflower receptacles. In this study, a new method was created for chemical space network analysis and classification of small samples, and applied to 104 compounds. Here, t-SNE (t-Distributed Stochastic Neighbor Embedding) dimensions were used to reduce coordinates as node locations and edge connections of chemical space networks, respectively, and molecules were grouped according to whether the edges were connected and the proximity of the node coordinates. Through detailed analysis of the structural characteristics and fingerprints of each classified group, our classification method attained good accuracy. Targets were then identified using reverse docking methods, and the active centers of the same types of compounds were determined by quantum chemical calculation. The results indicated that these compounds can be divided into nine groups, according to their mean within-group similarity (MWGS) values. The three families with the most members, i.e., the d-limonene group (18), α-pinene group (10), and γ-maaliene group (nine members) determined the protein targets, using PharmMapper. Structure fingerprint analysis was employed to predict the binding mode of the ligands of four families of the protein targets. Thence, quantum chemical calculations were applied to the active group of the representative compounds of the four families. This study provides further scientific information to support the use of sunflower receptacles.

Keywords: chemical space network; sunflower (*Helianthus annuus* L.); essential oils; fingerprint; clustering

1. Introduction

Sunflower (*Helianthus annuus* L.) belongs to the Compositae family (Asteraceae), which originated in South America and spread to China in the seventeenth century [1–3]. Sunflower has been widely cultivated in northeast China. After the seeds are used for oil extraction, sunflower receptacles have largely been discarded [4–6], which not only wastes resources, but also pollutes the environment. Sunflower receptacles contains many active compounds, including flavonoids [7,8], alkaloids [9,10], and chlorogenic acid [11,12]. However, there are few existing studies on the essential oil of sunflower receptacles. Therefore, it is in line with new pattern of green agriculture to change discarded sunflower receptacles from waste into a valuable resource.

Essential oils of sunflower are rich in unsaturated fatty acids, such as oleic and linoleic acids (ω-6), and are considered good for human health [13]. Essential oils of sunflower receptacles can be obtained by hydrodistillation [14].

In our previous studies, 101 compounds from the essential oils of sunflower (*Helianthus annuus* L.) receptacles were identified by gas chromatography-mass spectrometry (GC-MS) from three varieties of sunflowers, i.e., LD5009, SH363, and S606 [15,16]. The results

showed that eupatoriochromene may be one of the most important chemical compounds of sunflower receptacles for reducing uric acid. However, the functions of the other compounds of essential oils from sunflower receptacles remain unknown.

Chemical space network (CSN) techniques map chemical molecules into a visual space according to certain characteristics including molecular structure. CSN was initially designed as a coordinate-free threshold network using the Tanimoto coefficient as a continuous similarity measure [17]. However, it can only reflect the connection between molecules and not the relative distances of molecules in space. Several studies have proposed more novel CSNs, such as the TV-CSN [18] and Kamada–Kawai network [19], the latter attempting to introduce coordinates into the CSN. Compared to the traditional threshold CSN, in this study we added t-SNE dimensionality reduction to determine the spatial coordinates, to adapt to component clustering in mixtures with large differences in composition, such as essential oils.

Protein–ligand interaction fingerprints (IFPs) are binary one-dimensional representations of the three-dimensional structures of protein–ligand complexes, encoding the presence or absence of specific interactions between the binding pocket amino acids and the ligand. IFPs have successfully been applied for post-processing molecular docking results for G protein-coupled receptor (GPCR) ligand binding mode prediction and virtual ligand screening [20].

The purpose of this study was to compare the differences in the chemical compounds of essential oils from sunflower receptacles. Over 100 compounds were clustered by mapping into a CSN, and representative compounds from each group were selected. The target molecules of each group were identified using reverse docking to the group's representative compound. The active centers of the same type of compounds were determined by quantum chemical calculation. This study can provide reliable clues for the application of these compounds, and further scientific information to support the use of sunflower receptacles, which can reduce the waste of sunflower receptacles and increase the incomes of farmers.

2. Results

2.1. Cluster Analysis

The molecular similarities are shown for 104 compounds (Figure 1). Due to the complex composition of essential oils, similarity between molecules was found to be generally low, but the similarities between some molecules were obvious, as in the long-chain compounds (the upper left corner of the heatmap). A whitish area was observed in the heatmap (hierarchical clustering is shown in red), representing a large cluster of endocyclic compounds. Due to the special structure of ceertain molecules, their similarities with the other 103 molecules are all less than 0.45, so they did not participate in network generation. Finally, 91 of the 104 compounds had at least one edge with a similarity greater than 0.45, indicating their participation in the network generation.

The two-dimensional coordinates of 104 molecules calculated by t-SNE dimensionality reduction can be viewed in Table S2, and were used to help generate the CSN.

According to the edge data in Figure 1 and the node coordinate data in Table S2, we grouped molecules based on similarity greater than 0.45 and proximate node location. After filtering out groups whose total number of molecules was less than three, 73 of 91 compounds remained clustered into nine classes; see Figure 2 and Tables S3–S11. We can see that many related molecules were not grouped together because they were too far apart, which is exactly as expected after refining the groups. The α-pinene group, d-limonene group, α-muurolene group, and γ-maalineen group were more or less connected; traditional methods have difficulty distinguishing molecules that link multiple groups, but after introducing the coordinates obtained by dimensionality reduction, the four groups were visually separated. This offered preliminarily proof that our method is feasible for analysis of essential oils with large differences in components.
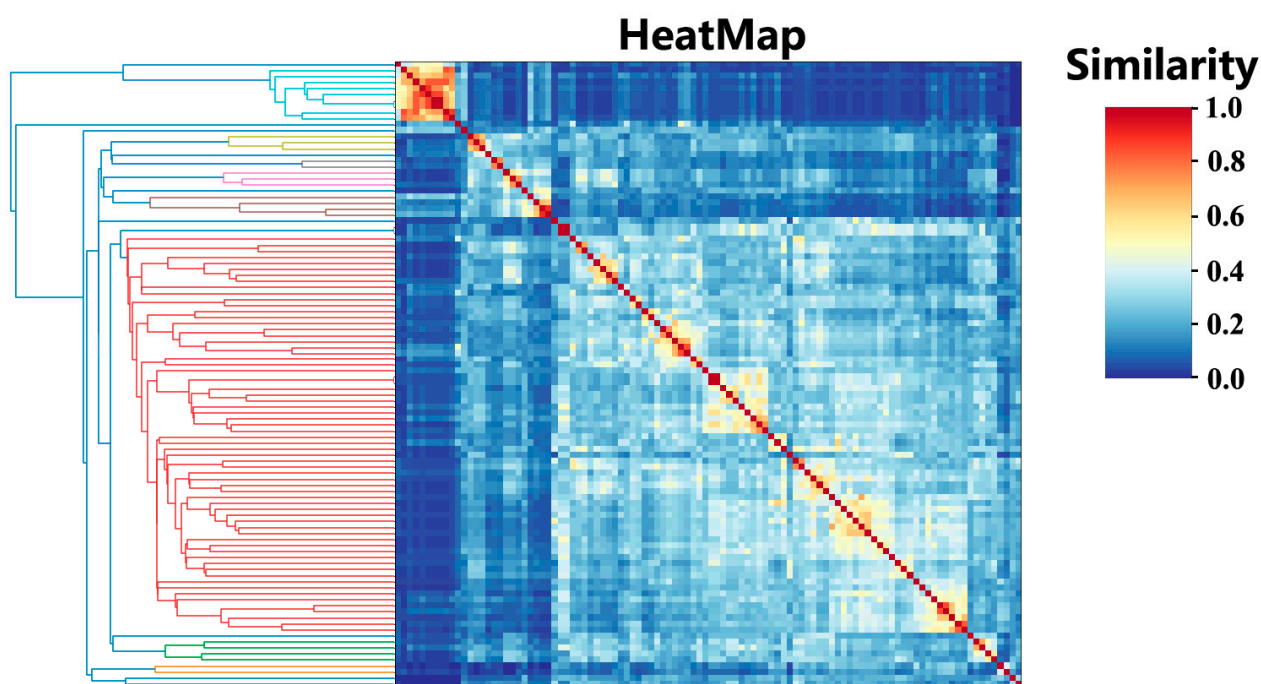
Figure 1. Similarity heatmap for 104 compounds. Where the similarity between molecules is greater, the color is red, and the lower the similarity, the bluer the color. The left part of the figure represents hierarchical clustering calculated according to the similarity values.

Figure 2. Chemical space network of 91 compounds. Nine groups with more than two molecules are colored, other groups with greater distance or less than three molecules are shown in gray.

Next, we analyzed the structural characteristics of each group classification. The linoleic acid group mainly comprised long chain compounds. The α-Pinene group compounds were found to be mainly b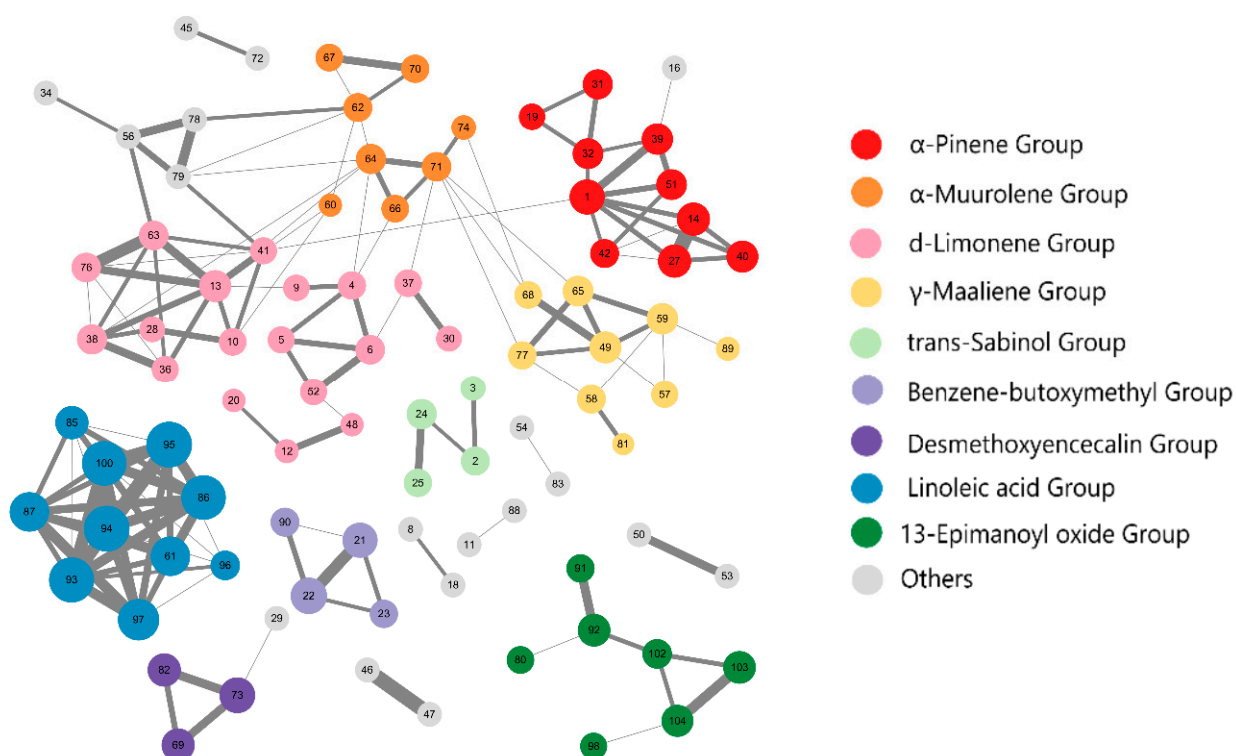icyclic monoterpenes and their oxygen-containing derivatives. Many structural types of bicyclic monoterpenes exist, including pinene, campene, carene, and others. Among these, the pinene and camphene types are the most stable. The first nine examples in Table S3 are pinene type, with a bridged ring skeleton of 2,6,6-trimethylbicyclo [3.1.1] heptane, pinene, rosinol, myrtenol, and verbenol. All are typical pinene-type compounds. The last example is a camphene type, with a bridged ring skeleton of 1,7,7-trimethylbicyclo [2.2.1] heptane. The camphene type compounds mostly exist as oxygen-containing derivatives, such as 6-camphenone. D-limonene group compounds are mainly monocyclic monoterpenes and their derivatives including formates, ketones, alcohols, except cis-Australinol, β-bisabolene, two monocyclic sesquiterpenes, and enols. γ-Maalineen group compounds are mainly tricyclic sesquiterpenes and their oxygen-containing derivatives, including acetates, lactones, and alcohols, except for trans-valerian terpene alcohol acetates, which are acetates of bicyclic sesquiterpenes. The compounds of epimanoly oxide group are mainly oxygen-containing derivatives of tricyclic or tetracyclic diterpenes, and include acids, alcohols, ketones, formate esters, and ethers. The group can be divided into four categories. The first category is tetracyclic diterpene with kaurine diterpenes as the core skeleton, like kauri aldehyde, kauri acid, H-Kauran-16-ol, and enantio–kaurane diterpenes. They are natural products with many important biological activities including antibacterial, anti-inflammatory, and anti-tumor effects [21]. The second category is ethyl isopimaric acid with tricyclic diterpene–pimarane diterpene as the core skeleton. The third category is ribenone and 13-epimanoyl oxide with tricyclic diterpenes–helichryllane diterpenes (sclareolide) as the core skeleton. Finally, the ricyclic-diterpenes and long-leaf aldehydes have more complex bridged ring structures. The compounds of the benzene–butoxymethyl group are alcohols, ethers, and dibutyl esters containing benzene rings. The compounds of trans-sabinol group compounds are bicyclic monoterpene compounds and their oxygen-containing derivatives, including alcohols and formate esters, all of which have bicyclic structure of 4-methyl-1-isopropylbicyclo [3.1.0] hexane. The compounds of desmethoxtencecalin share the structure of benzo-α-pyran, i.e., α-chromene.

In this study, the three families with the most members (the d-limonene (18), α-pinene (10), and γ-maaliene groups (nine) were designated the protein targets for further study. Although the linoleic acid group had 10 compounds, they have been found in many plants and well researched [22], so the linoleic acid group was not considered in the next study.

2.2. Reverse Docking, Structure Fingerprint, and Quantum Chemical Calculation Analysis
2.2.1. d-Limonene Group

There were 18 compounds in the d-limonene group, including d-limonene, the representative compound of the group (MWGS value 0.42), as shown in Table S3. The greater the MWGS value, the more closely a compound is related to other compounds in this family, and this was used to designate the predicted target protein in PharmMapper [23,24]. The predicted target was human placental estrone/DHEA sulfatase (ES, PDB ID is 1P49), which catalyzes the conversion of sulfated steroid precursors such as dehydroepiandrosterone sulfate (DHEA-S) and estrone sulfate to the free steroid [25]. Estrone sulfatase (ES) is one of the key enzymes involving in maintaining high levels of estrogen in breast tumor cells. The presence of ES in breast carcinomas has been related to breast cancer and X-linked ichthyosis, a disease of the skin [26]. Figure 3A shows the LUMO (lowest unoccupied molecular orbital) orbits of the d-limonen. LUMO is important for the establishment of the chemical bond and is integral in the sphere of spectroscopy. It depends on all coordinates of a system, providing a more efficient sampling method than a geometrical reaction coordinate, to better reflect the activities of the compound. It can be seen that the LUMO is concentrated on the propylene group, which will obtain electrons more easily and become chemically more active.
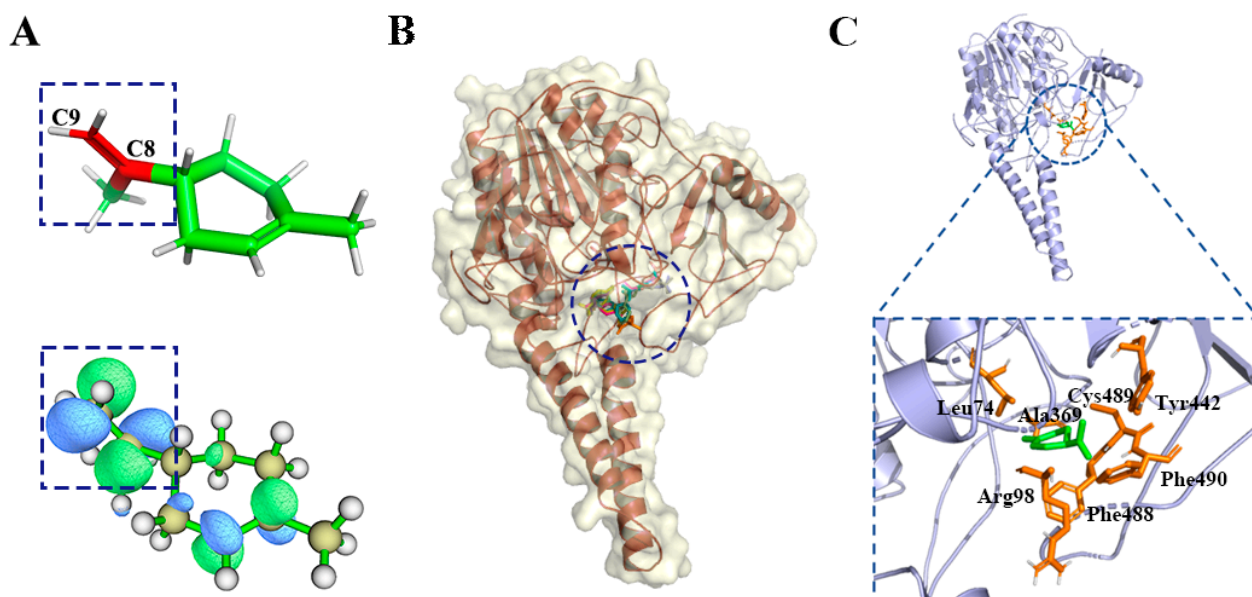
Figure 3. (**A**) LUMO orbit of d-limonene. (**B**) d-limonene group compounds docking with ES. (**C**) Active residues around d-limonene docking with ES.

Figure 3B,C show d-limonene group molecules binding in the active pocket of the ES, and d-limonene interacting with amino acid residues of ES. Subsequently. The interactions of the receptor to ligand can be determined in several contact types: Pi-orbital (PO), alkyl-pi (Ak), H-donor (HD), H-acceptor (HA), and sulfur bond (SF). Figure 4 revealed that L74, V101, V486, C489 and F488 interacted with most of the molecules in this group and hence may be important residues for ligand-binding to ES.

No.	P73:Ak	L74:Ak	R98:Ak	V101:Ak	L167:Ak	R168:Ak	V177:Ak	H290:PO	T291:HA	T291:HD	H346:HD	H346:PO	K368:HD	A369:Ak	A369:HA	C446:Ak	N447:HA	Y449:PO	H485:PO	V486:Ak	C487:HA	F488:PO	C489:Sf	C489:Ak	F490:PO
48	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0
9	0	1	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
5	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0
4	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0
12	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	1	1	0
13	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1	1
28	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
38	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
76	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1	0	1	0
52	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
36	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0
10	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
37	0	1	0	0	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0
41	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1
30	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0
63	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	1
6	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0
Frequency	0.06	0.72	0.17	0.50	0.17	0.17	0.39	0.22	0.06	0.06	0.11	0.39	0.11	0.44	0.11	0.33	0.06	0.11	0.50	0.83	0.17	0.61	0.11	0.50	0.11

Figure 4. Structure fingerprint of d-Limonene group. PO, pi-orbitals; Ak, alkyl; HD, H-donor; HA, H-acceptor; Sf, sulfur.

2.2.2. α-Pinene Group

There were 10 compounds in the α-pinene group. α-Pinene, the representative compound of the α-pinene group (MWGS value 0.47) was used for the predicted target protein in PharmMapper (Table S4). The most popular target protein of α-pinene group was vitamin D binding protein (DBP, PDB ID is 1J78) [27], which has many important functions, containing and transporting vitamin D3 metabolites, binding the globular actin, and transferring fatty acids to functions in the immune system.

Figure 5A shows the LUMO orbits of α-pinene. It can be seen that the LUMO was expressed at the C=C group, which can gain electrons more easily and become more

chemically active. Figure 5B,C shows the binding pose of ten compounds to DBP, and the representative compound of α-pinene group binding to the target protein. From Table 1, it can be seen that 10 compounds had interaction with the pi-orbital of F36 and alkyl-pi interaction with V88 and M107, respectively. Hence, F36, V88, and M107 played important roles in α-pinene group compounds' binding to DBP.
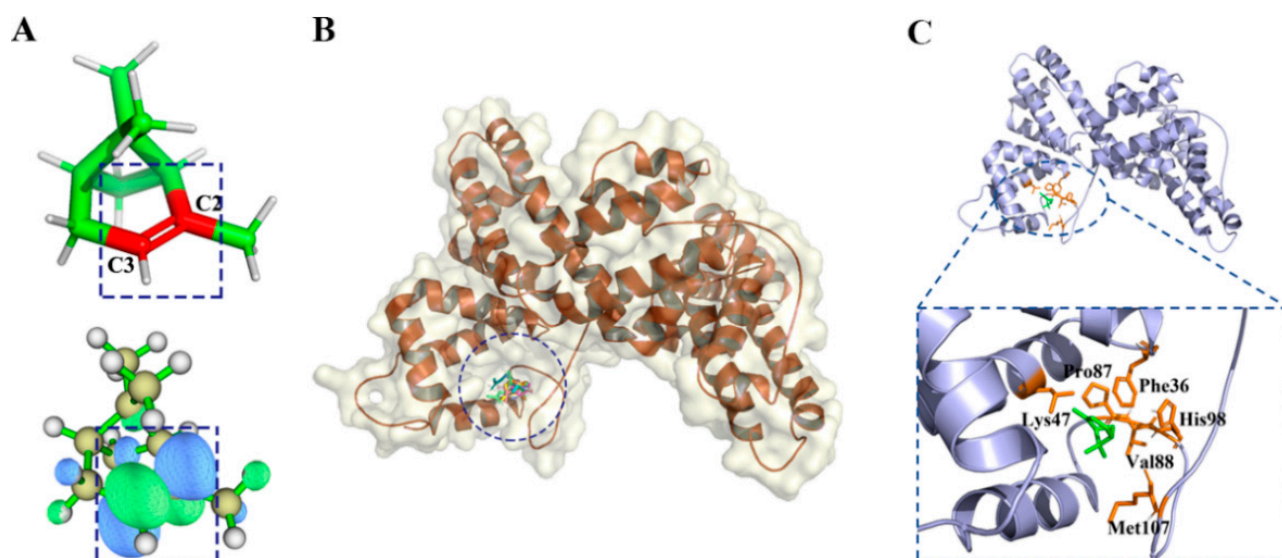

Figure 5. (**A**) LUMO orbit of α-Pinene. (**B**) α-Pinene group compounds docked to DBP. (**C**) Active residues around α-Pinene binding to DBP.

Table 1. Fingerprint of the α-pinene group.

No.	F36:PO	L47:Ak	S76:HD	S79:HD	P87:HA	P87:Ak	V88:Ak	H89:HD	H89:PO	M107:Ak	L110:Ak
51	1	0	0	1	1	0	1	0	1	1	0
19	1	1	0	0	0	0	1	1	0	1	0
31	1	0	0	0	0	0	1	0	1	1	1
14	1	1	0	0	0	1	1	0	1	1	0
39	1	0	0	1	0	0	1	0	1	1	0
40	1	0	0	1	0	0	1	0	1	1	0
42	1	1	1	0	0	1	1	0	1	1	0
27	1	0	0	1	0	0	1	0	1	1	1
32	1	0	0	0	0	0	1	0	1	1	1
1	1	1	0	0	0	1	1	0	1	1	0

PO, pi-orbitals; Ak, alkyl; HD, H-donor; HA, H-acceptor.

2.2.3. γ-Maaliene Group

There were nine compounds in the γ-maaliene group. γ-Maaliene, the representative compound of the γ-maaliene group (MWGS value is 0.42, see Table S6), was used as the ligand to predict target protein with PharmMapper. The most popular target protein of the γ-maaliene group was kinesin-like protein KIF11 (KSP, PDB ID is 2FKY), a motor protein required for establishing a bipolar spindle during mitosis [28]. KSP inhibitors have potential as general antiproliferative agents useful for the treatment of cancer.

Figure 6A shows LUMO orbits of the γ-maaliene. It can be seen that LUMO was expressed at the C=C group between C6 and C7, which can gain electrons more easily and become more chemically active. Figure 6B,C show the binding pose of nine compounds to KSP, and the γ-maaliene, the representative compound of the γ-maaliene group, binding to KIF11. Table 2 shows I36, P137, L214, and R218, which play important hydrophobic roles for compounds of γ-maaliene group.
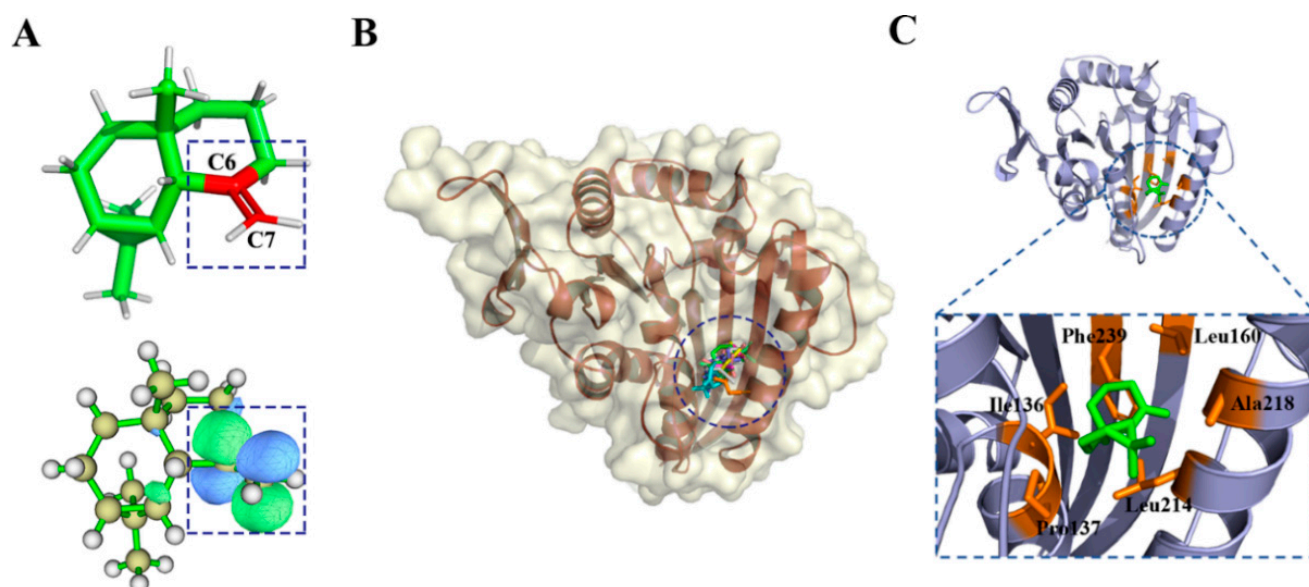
Figure 6. (**A**) LUMO orbit of γ-Maaliene; (**B**) γ-Maaliene group compounds docking with KIF11; (**C**) Active residues around γ-Maaliene binding to KIF11.

Table 2. Fingerprint of the γ-maaliene group.

No.	E116:HA	R119:Ak	I136:Ak	P137:Ak	L160:Ak	L172:Ak	Y211:PO	L214:HA	L214:Ak	E215:HD	A218:Ak
59	0	0	1	1	1	0	0	0	1	0	1
68	0	0	1	1	1	0	0	0	1	0	1
81	0	0	1	1	1	0	0	0	1	1	1
58	0	0	1	1	1	0	0	0	1	0	1
65	0	0	1	1	0	0	0	0	1	0	1
89	0	0	1	1	1	1	0	1	1	0	1
49	0	0	1	1	1	0	0	0	1	0	1
57	0	1	1	1	0	0	1	0	1	0	0
77	1	0	1	1	0	0	1	0	1	0	0

PO, pi-orbitals; Ak, alkyl; HD, H-donor; HA, H-acceptor.

3. Discussions

We tested the prepared bitters dataset (Table S13) from BitterDB (https://bitterdb.agri. huji.ac.il/dbbitter.php, accessed on 29 August 2022) using threshold CSN (THR CSN) and t-SNE dimensionality reduction CSN (t-SNE CSN), and results are shown in Figure 7. The Python script is available in the supplementary data. We observed that some molecules in the traditional THR CSNs were highly aggregated, but molecules with a generally high similarity in a large central area of the CSN were not subdivided. Molecules were well allocated to suitable locations in space according to their structural characteristics in t-SNE CSN. More importantly, the molecules in the center of the space were well clustered according to whether they were connected and how far apart they were positioned. We were then able to set the edge length threshold, to separate the group spatially.

t-SNE CSN is essentially a combination of two different representations (t-SNE and DiceSimilarity) of the 1024-bit Morgan fingerprint. Morgan fingerprinting is very suitable for terpenoids with complex topology in their essential oils, and our method can fully mine the information of the Morgan fingerprint. However, since t-SNE cannot deal with spatial discontinuity in two-dimensional space, wit is possible to obtain many molecules with higher similarity assigned to the opposite side of the CSN. This requires the design of algorithms, such as those incorporating machine learning.

Therefore, compared with THR CSN, t-SNE CSN has the advantage that a large number of aggregated molecules can be further subdivided after the introduction of coordinates, which helps to unearth more compound groups.
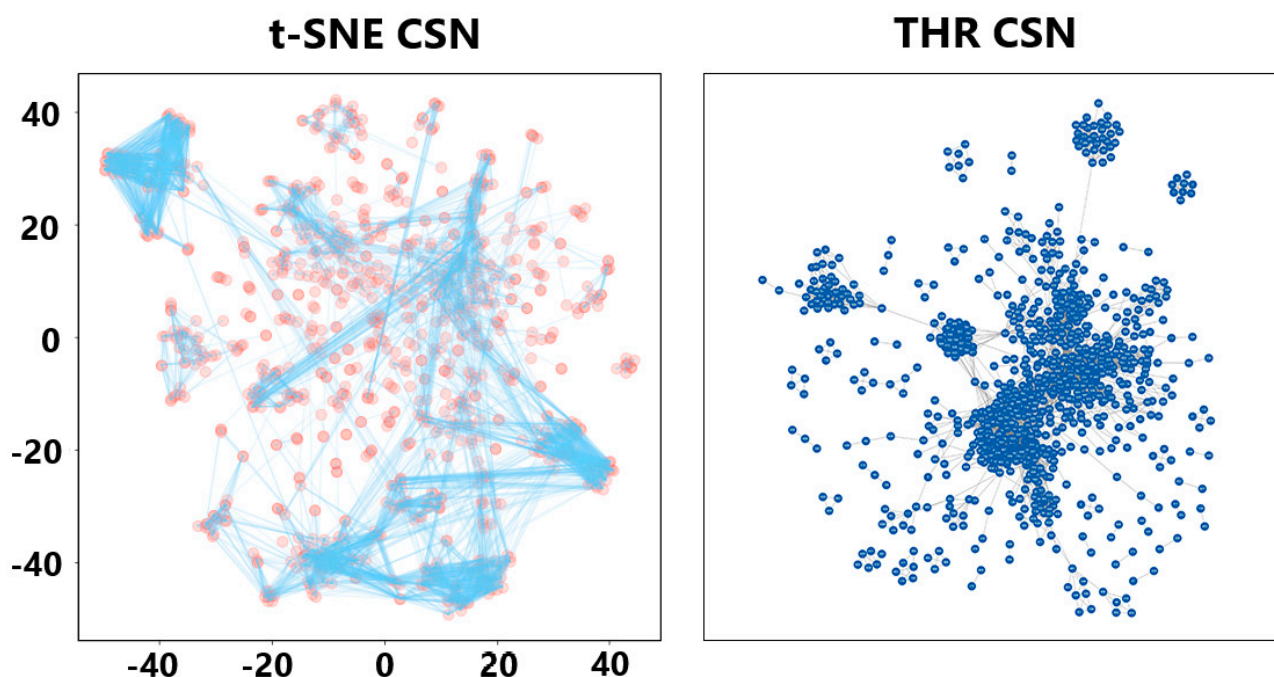

Figure 7. t-SNE CSN and THR CSN for the bitters dataset. Molecules in the t-SNE CSN are represented by red circles, and edges with a similarity greater than 0.45 are indicated in blue. In the THR CSN, molecules and edges are represented as blue circles and black lines, respectively.

4. Materials and Methods

4.1. Cluster Analysis

Firstly, we determined the CSN edges. The SMILES structural formulas of 104 compounds (three isomers added) were queried on PubChem, based on the compound names inferred from the peak time of GC-MS [15,16]. The Morgan fingerprints of the molecules [29] were extracted using Rdkit [30–32], where the fingerprint radius was set to two. Fingerprint similarity uses Dice coefficient, and the calculation formula is as follows:

$$\text{DiceSimilarity}(a, b) = \frac{2 \times (a \cap b)}{a + b} \tag{1}$$

where a and b are the substructure features of the two molecules, respectively.

We calculated fingerprint similarity among 104 compounds and generated hierarchical clustering and correlation heatmaps with scikit-learn [33,34]. According to the similarity data, the similarity threshold was set to 0.45. The similarity of two molecules greater than the threshold forms an edge in the chemical space network, but the relative position of the compounds cannot be determined at this stage.

Next, we determined the node coordinates of CSN. Therefore, t-SNE [35,36] was used to reduce the dimensionality of the Morgan fingerprints of 104 compounds in order to obtain their relative spatial positions. To unify the number of bits of the fingerprint vector, the ECFP fingerprint [37] was generated using the explicit bitvectors method during dimension reduction; the number of bits was set to 1024, and the radius was two.

Then, we built a chemical space network, constructed according to the node coordinates and edge connections generated in the previous two steps.

Finally, we clustered and identified the representative compounds for each group. According to whether there were edge connections between molecules and whether the

positions were close, nine groups of compounds were determined manually. The mean within-group similarity (MWGS) was calculated for the molecules of each group:

$$\text{MWGS}(S_i, n) = \frac{(\sum_{i=1}^{n} S_i) - 1}{n} \tag{2}$$

where n is the number of molecules of the group, and is the similarity between the molecule and the i-th molecule in the group.

The compound with the largest MWGS in the group was considered the representative compound of the group.

In short, 91 of the 104 compounds participated in the generation of the network, and finally 73 of them were formed into nine groups, according to edge information and coordinate information (see Figure 8).
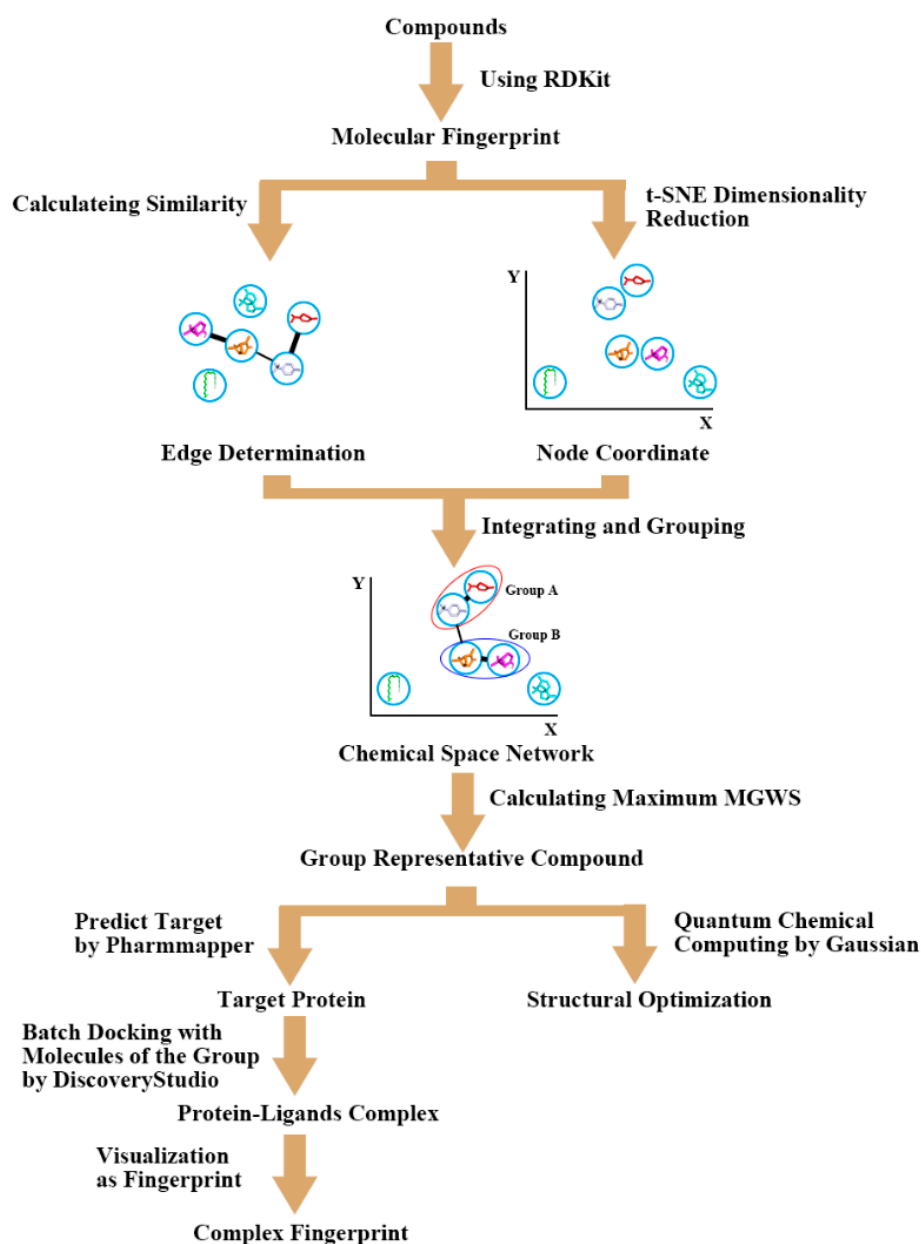

Figure 8. The workflow of the study methods.

4.2. Group Docking and Structure Fingerprint Analysis

Using OpenBabel [38,39] to convert the SMILES formula of each group of representative compounds into mol2 format, we returned to PharmMapper [23,24] to predict the target. Then, we applied Biovia Discovery Studio to dock all molecules of each group with the corresponding target proteins. Molecular fingerprints were extracted using a Python script based on the docking results from Discovery Studio.

4.3. Quantum Chemical Calculations

The quantum chemical calculations were carried out using the B3LYP function [40–43] implemented in the Gaussian 09 program at the 6–31 G* set [44,45]. Frequency calculations were performed to obtain free energy corrections at 298.15 K and 1atm pressure. Multiwfn [46,47], a multifunctional program for wave function analysis of quantum chemical calculation results, was used to analyze the weak interaction of the ligands. The number of grids was set to $200 \times 200 \times 200$ in three-dimensional space.

The 5000 frames of trajectories were extracted to average the density. To analyze traditional H-bond occupancy, the angle and distance between the donor and acceptor were set to 35° and 3.5 Å, respectively.

5. Conclusions

In this study, 104 compounds from essential oil in sunflower receptacles were mapped and grouped in our designed chemical space network (t-SNE CSN). The results indicated that these compounds can be divided into nine groups according to their MWGS value. PharmMapper was utilized to identify the target protein of the three families with the most members, i.e., the d-limonene, α-pinene, and γ-maaliene groups. The binding modes of the ligands of the three families to the target protein were indicated using structure fingerprint analysis. The active center of the same type of compounds was determined by quantum chemical calculation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms231710169/s1.

Author Contributions: Conceptualization, Y.H.; methodology, Y.H.; software, K.L.; validation, Y.H. and L.H.; formal analysis, Y.H.; investigation, Y.H.; resources, Y.H.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H.; visualization, Y.H.; supervision, Y.H.; project administration, W.H.; funding acquisition, W.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: MDPI Research Data Policies.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zamani, S.; Naderi, M.R.; Soleymani, A.; Nasiri, B.M. Sunflower (*Helianthus annuus* L.) biochemical properties and seed components affected by potassium fertilization under drought conditions. *Ecotoxicol. Environ. Saf.* **2020**, *190*, 110017. [CrossRef]
2. Radonic, L.M.; Lewi, D.M.; López, N.E.; Hopp, H.E.; Escandón, A.S.; Bilbao, M.L. Sunflower (*Helianthus annuus* L.). *Methods Mol. Biol.* **2015**, *1224*, 47–55. [CrossRef]
3. Lewi, D.M.; Hopp, H.E.; Escandón, A.S. Sunflower (*Helianthus annuus* L.). *Methods Mol. Biol.* **2006**, *343*, 291–297. [CrossRef]
4. Smith, B.D. Eastern North America as an independent center of plant domestication. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12223–12228. [CrossRef] [PubMed]
5. Lawson, S.K.; Sharp, L.G.; Powers, C.N.; McFeeters, R.L.; Satyal, P.; Setzer, W.N. Essential Oil Compositions and Antifungal Activity of Sunflower (Helianthus) Species Growing in North Alabama. *Appl. Sci.* **2019**, *9*, 3179. [CrossRef]
6. Shi, B.; Zhao, J. Recent progress on sunflower broomrape research in China. *OCL* **2020**, *27*, 30. [CrossRef]

7.   Serafini, M.; Peluso, I.; Raguzzini, A. Flavonoids as anti-inflammatory agents. *Proc. Nutr. Soc.* **2010**, *69*, 273–278. [CrossRef]
8.   Wen, K.; Fang, X.; Yang, J.; Yao, Y.; Nandakumar, K.S.; Salem, M.L.; Cheng, K. Recent Research on Flavonoids and their Biomedical Applications. *Curr. Med. Chem.* **2021**, *28*, 1042–1066. [CrossRef] [PubMed]
9.   Cinelli, M.A.; Jones, A.D. Alkaloids of the Genus Datura: Review of a Rich Resource for Natural Product Discovery. *Molecules* **2021**, *26*, 2629. [CrossRef]
10.   Bhambhani, S.; Kondhare, K.R.; Giri, A.P. Diversity in Chemical Structures and Biological Properties of Plant Alkaloids. *Molecules* **2021**, *26*, 3374. [CrossRef]
11.   Nabavi, S.F.; Tejada, S.; Setzer, W.N.; Gortzi, O.; Sureda, A.; Braidy, N.; Daglia, M.; Manayi, A.; Nabavi, S.M. Chlorogenic Acid and Mental Diseases: From Chemistry to Medicine. *Curr. Neuropharmacol.* **2017**, *15*, 471–479. [CrossRef] [PubMed]
12.   Miao, M.; Xiang, L. Pharmacological action and potential targets of chlorogenic acid. *Adv. Pharmacol.* **2020**, *87*, 71–88. [CrossRef] [PubMed]
13.   Galúcio, C.S.; Souza, R.A.; Stahl, M.A.; Sbaite, P.; Benites, C.I.; Maciel, M.R.W. Physicochemical characterization of monoacylglycerols from sunflower oil. *Procedia Food Sci.* **2011**, *1*, 1459–1464. [CrossRef]
14.   Aguirre, M.; Velasco, J.; Ruiz-Méndez, M.V. Characterization of sunflower oils obtained separately by pressing and subsequent solvent extraction from a new line of seeds rich in phytosterols and conventional seeds. *OCL—Ol. Corps Gras Lipides* **2014**, *21*, 5. [CrossRef]
15.   Liu, X.S.; Gao, B.; Dong, Z.D.; Qiao, Z.A.; Yan, M.; Han, W.W.; Li, W.N.; Han, L. Chemical Compounds, Antioxidant Activities, and Inhibitory Activities Against Xanthine Oxidase of the Essential Oils From the Three Varieties of Sunflower (*Helianthus annuus* L.) Receptacles. *Front. Nutr.* **2021**, *8*, 737157. [CrossRef] [PubMed]
16.   Liu, X.S.; Gao, B.; Li, X.L.; Li, W.N.; Qiao, Z.A.; Han, L. Chemical Composition and Antimicrobial and Antioxidant Activities of Essential Oil of Sunflower (*Helianthus annuus* L.) Receptacle. *Molecules* **2020**, *25*, 5244. [CrossRef]
17.   Zhang, B.; Vogt, M.; Maggiora, G.M.; Bajorath, J. Comparison of bioactive chemical space networks generated using substructure- and fingerprint-based measures of molecular similarity. *J. Comput. Aided Mol. Des.* **2015**, *29*, 595–608. [CrossRef]
18.   Wu, M.; Vogt, M.; Maggiora, G.M.; Bajorath, J. Design of chemical space networks on the basis of Tversky similarity. *J. Comput. Aided Mol. Des.* **2016**, *30*, 1–12. [CrossRef]
19.   Vega de León, A.; Bajorath, J. Design of chemical space networks incorporating compound distance relationships. *F1000Research* **2016**, *5*, 2634. [CrossRef]
20.   Vass, M.; Kooistra, A.J.; Ritschel, T.; Leurs, R.; de Esch, I.J.; de Graaf, C. Molecular interaction fingerprint approaches for GPCR drug discovery. *Curr. Opin. Pharmacol.* **2016**, *30*, 59–68. [CrossRef]
21.   De las Heras, B.; Hoult, J.R. Non-cytotoxic inhibition of macrophage eicosanoid biosynthesis and effects on leukocyte functions and reactive oxygen species of two novel anti-inflammatory plant diterpenoids. *Planta Med.* **1994**, *60*, 501–506. [CrossRef] [PubMed]
22.   Koba, K.; Yanagita, T. Health benefits of conjugated linoleic acid (CLA). *Obes. Res. Clin. Pract.* **2014**, *8*, e525–e532. [CrossRef]
23.   Wang, X.; Shen, Y.; Wang, S.; Li, S.; Zhang, W.; Liu, X.; Lai, L.; Pei, J.; Li, H. PharmMapper 2017 update: A web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res.* **2017**, *45*, W356–W360. [CrossRef] [PubMed]
24.   Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38*, W609–W614. [CrossRef]
25.   Hernandez-Guzman, F.G.; Higashiyama, T.; Pangborn, W.; Osawa, Y.; Ghosh, D. Structure of human estrone sulfatase suggests functional roles of membrane association. *J. Biol. Chem.* **2003**, *278*, 22989–22997. [CrossRef] [PubMed]
26.   Ahmed, S.; Owen, C.P.; James, K.; Sampson, L.; Patel, C.K. Review of estrone sulfatase and its inhibitors–an important new target against hormone dependent breast cancer. *Curr. Med. Chem.* **2002**, *9*, 263–273. [CrossRef] [PubMed]
27.   Verboven, C.; Rabijns, A.; De Maeyer, M.; Van Baelen, H.; Bouillon, R.; De Ranter, C. A structural basis for the unique binding features of the human vitamin D-binding protein. *Nat. Struct. Biol.* **2002**, *9*, 131–136. [CrossRef] [PubMed]
28.   Fraley, M.E.; Garbaccio, R.M.; Arrington, K.L.; Hoffman, W.F.; Tasber, E.S.; Coleman, P.J.; Buser, C.A.; Walsh, E.S.; Hamilton, K.; Fernandes, C.; et al. Kinesin spindle protein (KSP) inhibitors. Part 2: The design, synthesis, and characterization of 2,4-diaryl-2,5-dihydropyrrole inhibitors of the mitotic kinesin KSP. *Bioorganic Med. Chem. Lett.* **2006**, *16*, 1775–1779. [CrossRef]
29.   Martin, L.J.; Bowen, M.T. Comparing Fingerprints for Ligand-Based Virtual Screening: A Fast and Scalable Approach for Unbiased Evaluation. *J. Chem. Inf. Model.* **2020**, *60*, 4536–4545. [CrossRef]
30.   Lovrić, M.; Molero, J.M.; Kern, R. PySpark and RDKit: Moving towards Big Data in Cheminformatics. *Mol. Inf.* **2019**, *38*, e1800082. [CrossRef]
31.   Coley, C.W.; Green, W.H.; Jensen, K.F. RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application. *J. Chem. Inf. Model.* **2019**, *59*, 2529–2537. [CrossRef] [PubMed]
32.   Kruger, F.; Stiefl, N.; Landrum, G.A. rdScaffoldNetwork: The Scaffold Network Implementation in RDKit. *J. Chem. Inf. Model.* **2020**, *60*, 3331–3335. [CrossRef]
33.   Bac, J.; Mirkes, E.M.; Gorban, A.N.; Tyukin, I.; Zinovyev, A. Scikit-Dimension: A Python Package for Intrinsic Dimension Estimation. *Entropy* **2021**, *23*, 1368. [CrossRef]
34.   Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830.

35. Linderman, G.C.; Steinerberger, S. Clustering with t-SNE, provably. *SIAM J. Math. Data Sci.* **2019**, *1*, 313–332. [CrossRef] [PubMed]

36. Kang, B.; García García, D.; Lijffijt, J.; Santos-Rodríguez, R.; De Bie, T. Conditional t-SNE: More informative t-SNE embeddings. *Mach. Learn.* **2021**, *110*, 2905–2940. [CrossRef]

37. Le, T.; Winter, R.; Noé, F.; Clevert, D.A. Neuraldecipher—Reverse-engineering extended-connectivity fingerprints (ECFPs) to their molecular structures. *Chem. Sci.* **2020**, *11*, 10378–10389. [CrossRef]

38. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]

39. O'Boyle, N.M.; Morley, C.; Hutchison, G.R. Pybel: A Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* **2008**, *2*, 5. [CrossRef]

40. Lehtola, S.; Dimitrova, M.; Fliegl, H.; Sundholm, D. Benchmarking Magnetizabilities with Recent Density Functionals. *J. Chem. Theory Comput.* **2021**, *17*, 1457–1468. [CrossRef]

41. Sarkar, R.; Boggio-Pasqua, M.; Loos, P.F.; Jacquemin, D. Benchmarking TD-DFT and Wave Function Methods for Oscillator Strengths and Excited-State Dipole Moments. *J. Chem. Theory Comput.* **2021**, *17*, 1117–1132. [CrossRef] [PubMed]

42. Kanungo, B.; Zimmerman, P.M.; Gavini, V. A Comparison of Exact and Model Exchange-Correlation Potentials for Molecules. *J. Phys. Chem. Lett.* **2021**, *12*, 12012–12019. [CrossRef]

43. Han, W.; Zhu, J.; Wang, S.; Xu, D. Understanding the Phosphorylation Mechanism by Using Quantum Chemical Calculations and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2017**, *121*, 3565–3573. [CrossRef] [PubMed]

44. Samuel, Y.; Garg, A.; Mulugeta, E. Synthesis, DFT Analysis, and Evaluation of Antibacterial and Antioxidant Activities of Sulfathiazole Derivatives Combined with In Silico Molecular Docking and ADMET Predictions. *Biochem. Res. Int.* **2021**, *2021*, 7534561. [CrossRef]

45. Mora, J.R.; Kirby, A.J.; Nome, F. Theoretical study of the importance of the spectator groups on the hydrolysis of phosphate triesters. *J. Org. Chem.* **2012**, *77*, 7061–7070. [CrossRef]

46. Zhang, J.; Lu, T. Efficient evaluation of electrostatic potential with computerized optimized code. *Phys. Chem. Chem. Phys.* **2021**, *23*, 20323–20328. [CrossRef]

47. Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592. [CrossRef] [PubMed]