processes

MDPI

Article

Evaluation of the Lipophilicity of 3,28-Disubstituted Betulin Derivatives with Promising Biological Properties

Elwira Chrobak [1,*], Katarzyna Bober-Majnusz [2,*], Marta Świtalska [3], Joanna Wietrzyk [3] and Ewa Bębenek [1]

[1] Department of Organic Chemistry, Faculty of Pharmaceutical Sciences in Sosnowiec, Medical University of Silesia in Katowice, 4 Jagiellońska, 41-200 Sosnowiec, Poland; ebebenek@sum.edu.pl
[2] Department of Analytical Chemistry, Faculty of Pharmaceutical Sciences in Sosnowiec, Medical University of Silesia in Katowice, 41-200 Sosnowiec, Poland
[3] Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, 12 Rudolfa Weigla Str., 53-114 Wrocław, Poland; marta.switalska@hirszfeld.pl (M.Ś.); joanna.wietrzyk@hirszfeld.pl (J.W.)
* Correspondence: echrobak@sum.edu.pl (E.C.); bober@sum.edu.pl (K.B.-M.); Tel.: +48-32-364-16-66 (E.C.)

Abstract

The identification of bioactive substances among new chemical compounds is based on analyzing the relationships between structure, physicochemical properties, and potential biological activity. The aim of this study was to characterize the physicochemical properties of a group of 3,28-disubstituted betulin derivatives, which have demonstrated promising antiproliferative activity in an in vitro study. The experimental lipophilicity parameters of betulin derivatives were obtained by RP-TLC and compared with theoretical values determined using various computational programs. Physicochemical and pharmacokinetic parameters were calculated using the SwissADME and pkCSM programs. The relationships between lipophilicity parameters (R_{M0} and $logP_{TLC}$) and the anticancer activity, and physicochemical and pharmacokinetic parameters of the studied triterpenoids were analyzed. Chemometric analysis (cluster analysis, principal component analysis, and the sum of ranking differences analysis) was performed. Significant correlations were demonstrated between R_{M0} and $logP_{TLC}$, as well as theoretically determined lipophilicity values, in the tested group of compounds. Propynoyl derivatives **4a**, **5a** and **6a** with high antiproliferative activity against MV4-11, PC-3 and Hs249T cells are characterized by higher lipophilicity than their hydroxyl analogs (compounds **4**, **5** and **6**).

Keywords: betulin derivatives; lipophilicity; RP-TLC method

Academic Editor: Ibrahim Abu-Reidah

Received: 5 January 2026
Revised: 16 February 2026
Accepted: 18 February 2026
Published: 20 February 2026

1. Introduction

Discovering a new drug has always been a costly and lengthy process. Researching and developing bioactive substances with the right properties involves significant risk, investment, and time. Understanding the physicochemical properties of a drug candidate compound can increase the likelihood of effective delivery and therapeutic success. Previous drug discovery approaches were based on the assumption that greater in vitro potency would lead to more effective therapy. Today, many pharmaceutical companies focus on analyzing compound properties early in preclinical development to reduce the costs of potential failures in later stages [1,2].

Natural products often exhibit more favorable physicochemical property profiles compared to drugs and are comparable to compounds obtained synthetically via ombinatorial chemistry [2]. Despite these advantages, classical processes for isolating and identifying new compounds from natural sources are insufficiently efficient compared to other drug

discovery methods used by pharmaceutical companies. To mor effectively exploit the properties of naturally occurring substances, they are considered as lead structures in the creation of libraries of derivatives based on natural products [3].

Betulin and betulinic acid occur in plants belonging to various families, e.g., birch (*Betulaceae*), rose (*Rosaceae*), plane (*Platanaceae*), beech (*Fagaceae*), and buckthorn (*Rhamnaceae*). Betulinic acid, which occurs in very small amounts in natural resources, can be synthesized by the oxidation of the primary hydroxyl group at the C-28 position of betulin [1,4,5]. Betulin and most of its derivatives are characterized by low water solubility and high lipophilicity. These characteristics play an important role in the interpretation of biological activity studies conducted in cell cultures [6]. The high level of safety of betulin and betulinic acid (low toxicity) has been a significant argument for conducting intensive work on improving their solubility to achieve stronger and more effective action [7]. To this end, the focus was primarily on functionalizing the alcohol groups at C-3 and C-28 in betulin or the acid group at C-17 in betulinic acid, creating esters, amides, carbamates, sulfates, sulfobetaines, and phosphates. Some of the obtained betulin and betulinic acid derivatives, such as uracil esters, glycerol esters, saponins, and amides conjugated with amino acids, demonstrated increased solubility compared to their precursors. Such structural changes were intended to improve pharmacokinetic properties [1,8–11].

The identification of bioactive substances among new chemical compounds is closely linked to the knowledge of the relationships between structure, physicochemical properties, and the potential activities that a new molecule may exhibit in a biological environment. Analyzing various series of derivatives, it can be concluded that the nature of the substituent attached to the starting structure of the molecule has a significant impact on the activity of the resulting derivatives. The early phase of drug design primarily involves determining the impact of structural changes on the corresponding properties of new biologically active molecules, most often based on structure–activity relationship models such as SAR (Structure–Activity Relationship) and QSAR (Quantitative Structure–Activity Relationship) [2]. QSAR studies utilize the chemical characterization of a molecule in numerical form using so-called molecular descriptors. Structural descriptors reflect the molecular composition of a compound with no information about its topology, i.e., the number of atoms, number of bonds, atom type, number of rings, and molecular weight (MW). Topological descriptors represent the interconnections of atoms within molecules. Electronic descriptors (dipole moment, energy of HOMO and LUMO orbitals) are used to describe the binding properties of molecules or atoms, as well as molecular fragments. Geometric descriptors, calculated from the 3D coordinates of atoms, contain information about the size, shape, and spatial arrangement. Thermodynamic descriptors link chemical structure to chemical activity. Using descriptors to determine the molecular structure of biologically active compounds is a fundamental method for discovering new lead structures [12].

Among the numerous molecular descriptors used in QSAR studies, lipophilicity plays a significant role as the most important physicochemical parameter of a molecule, significantly influencing its bioactivity. This parameter largely determines the transport of substances across the cell membrane, influencing the formation of complexes between tested substances and plasma proteins or the appropriate receptor at the site of action. Therefore, assessing the lipophilic nature of compounds is an essential step in the modern drug design process. RP-TLC (reversed-phase thin-layer chromatography) is a convenient method for lipophilicity testing due to its simplicity, low instrumentation requirements, the use of small amounts of substance, accuracy and reproducibility. It also provides the possibility of determining lipophilicity between structurally similar molecules. This method involves the use of two phases: a stationary and a mobile phase. The nonpolar and hydrophobic stationary phase involves silica gel modified by the addition of long carbon

chains (e.g., C2, C8, and C18). The most commonly used gel is octadecylsilane (RP-18), which has 18 carbon atoms. The mobile phase is a polar solvent system, most often an aqueous solution of acetone or methanol; other solvents such as ethanol, tetrahydrofuran, dioxane, or acetonitrile are also used [13–15].

This work continues research on the synthesis and properties of betulin and betulinic acid derivatives. Many structural modifications of terpenoids have been described in the literature, but the combination of an alkynyl and a carboxyacyl moiety in a single molecule is a novelty. Therefore, we deemed it necessary to further characterize these compounds. Previously synthesized compounds with the structure of 28-alkynyl derivatives of 3-carboxyacylbetuline were tested in vitro for their biological activity [16]. The promising results obtained regarding their antiproliferative activity prompted us to characterize their other properties relevant to their potential use as drugs. One of the parameters that significantly influences the in vivo fate of a chemical compound used as a drug is lipophilicity.

The conducted studies allowed us to determine the lipophilicity parameters of the title compounds (Table 1) using the RP-TLC method and compare the experimental and theoretical results obtained using various computational methods.

Table 1. Structural formulae of the tested compounds.

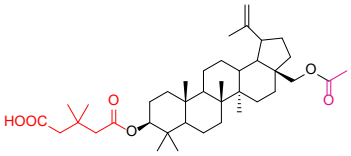
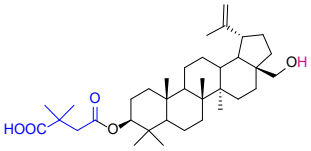
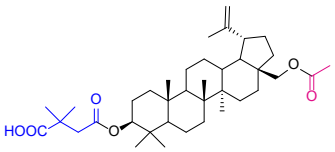
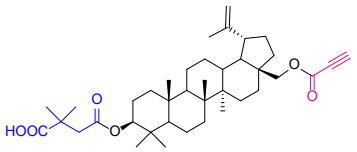
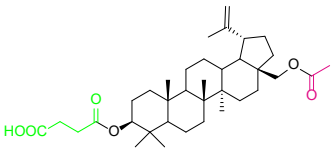
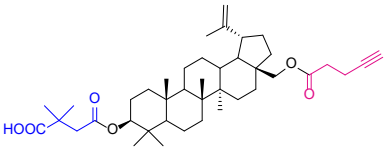
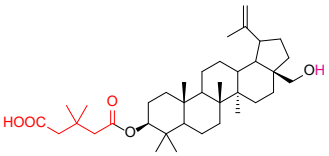
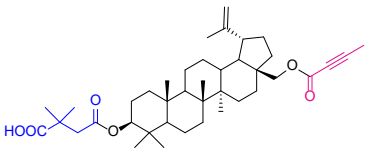
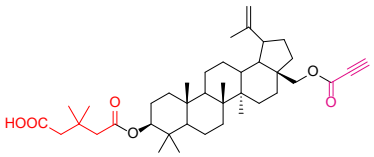
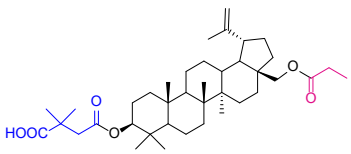
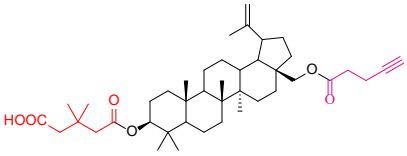
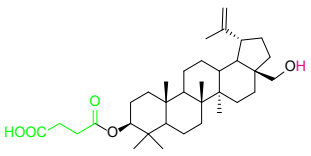
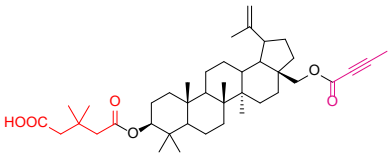
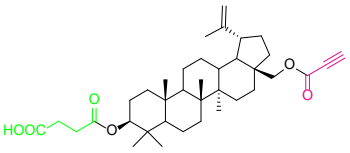
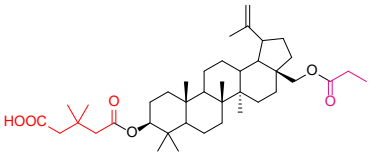
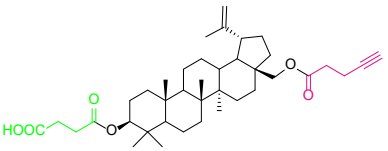
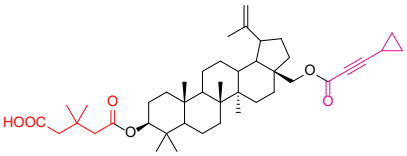
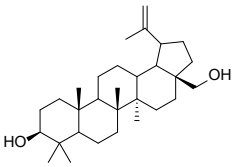
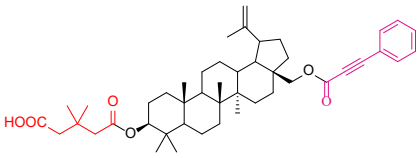
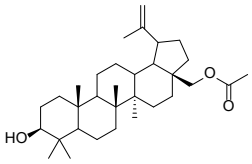
No. of Compound	Chemical Structure	No. of Compound	Chemical Structure
1		**5**	
2		**5a**	
3		**5b**	
4		**5c**	
4a		**5d**	
4b		**6**	

Table 1. *Cont.*

No. of Compound	Chemical Structure	No. of Compound	Chemical Structure
4c		**6a**	
4d		**6b**	
4e		**betulin (BET)**	
4f		**28-acetylbetulin (28-BET)**	

As part of the conducted work, we decided to analyze the relationship between lipophilicity parameters (R_{M0} and $logP_{TLC}$) and the anticancer activity, physicochemical, and pharmacokinetic parameters of the studied triterpenoids. Furthermore, chemometric analysis was performed, including CA (cluster analysis), PCA (principal component analysis), and SRD (sum of ranking differences).

2. Materials and Methods

2.1. Chemicals

In lipophilicity studies performed by reversed-phase thin-layer chromatography (RP-TLC), acetone (Eurochem BGD, Tarnów, Poland) and TRIS buffer with pH 7.4 (Merck, Darmstadt, Germany) were used to prepare the mobile phase. All standard compounds [acetanilide, prednisone, benzophenone, testosterone, anthracene, dibenzyl, 9-phenylanthracene, and DDT (dichlorodiphenyltrichloroethane) used to prepare the standard curve were purchased from Merck (Darmstadt, Germany). The studied betulin derivatives **1**–**6**, **4a**–**4f**, **5a**–**5d** and **6a**–**6b** were obtained via the synthetic route described previously in the literature [16].

2.2. Chromatography Procedure

RP-TLC used silica gel 60 RP-18 F_{254S} plates with dimensions of 5 cm × 10 cm (Merck, Darmstadt, Germany). All the tested compounds were dissolved in chloroform (concentration 2 mg/mL). Chromatograms were developed in chromatography chambers saturated with vapors of a mixture of acetone (Merck, Darmstadt, Germany) and 0.2 M Tris buffer (aqueous solution of tris-hydroxymethyl)aminomethane (Merck, Darmstadt, Germany) at pH 7.4. Six chromatography chambers were prepared with acetone content ranging from 65 to 90% (in 5% increments). Three chromatograms were performed for each compound in a given system. The chromatograms of betulin and its derivatives were visualized by

spraying with a 10% solution of sulfuric acid in ethanol and then heated to 110 °C. Spots of standard substances were observed under UV light at 254 nm.

The value of retardation factor (R_F) was determined for the chromatographic spots obtained. The R_F was converted into R_M, according to the following equation:

$$R_M = \log(\frac{1}{R_F} - 1) \quad (1)$$

R_M was linearly dependent on the organic part of mobile phase (acetone) and the R_{M0} value was obtained by extrapolating its concentration to zero, according to the next equation:

$$R_M = R_{M0} = bC \quad (2)$$

where C is the concentration of organic part (acetone) in the mobile phase and b is the regression term.

2.3. Methods of Chemometric Analysis

2.3.1. Correlation Analysis

Correlation analysis is a statistical method that measures the relationship between two or more variables. The results are presented as a correlation coefficient, whose absolute value indicates the strength of the relationship. A correlation is considered statistically significant if the significance level is equal to or less than 0.05.

2.3.2. Cluster Analysis (CA)

Cluster analysis is a technique used to group objects with similar characteristics. This analysis reduces the dimensionality of the similarity matrix by calculating the distances between objects. As a result, all objects are organized into a single cluster that contains smaller sub-clusters. Conclusions about the similarity of objects in a particular cluster are based on the analyst's experience and ability to identify common features.

2.3.3. Principal Component Analysis (PCA)

The above method is a statistical technique used for dimensionality reduction. It means that a set of multiple variables is transformed into a new set of uncorrelated variables. These new variables are known as principal components. Each subsequent principal component explains as much of the variability in the data as possible. The number of principal components is selected based on the Kaiser criterion or a scree plot. The selected principal components should explain as much of the variability in the data as possible. Very often, data standardization is necessary before PCA, due to the different scales of the variables. Standardization means expressing them in units of standard deviation values.

2.3.4. Sum of Ranking Differences (SRD)

The sum of ranking differences is one of the nonparametric methods that allows us to estimate the quality of the created models describing the system [17]. This statistical approach is valuable in chemometrics [18–20]. For each data point, we calculated the difference between its value and the data set mean. Each of these differences is squared. Then, we summed all these squares, creating a single number that determines the total variability of the data. Summarizing the sum of squares is a measure of the variability of the data, and its use allows us to assess how well the model fits the observations. Low values of the sum of squared differences indicate that the statistical model fits the data well. High values of the sum of squared differences suggest that the model fits the data poorly, and the predictions are subject to significant errors.

Correlation, cluster analysis and principal component analysis were performed via the use of STATISTICA 13.3 software, whilst the sum of ranking differences analysis was performed via Excel.

3. Results and Discussion

Chromatographic analysis was used to determine the values of the R_{M0} parameter for standard substances, including acetanilide, prednisone, benzophenone, testosterone, anthracene, dibenzyl, 9-phenylanthracene, and DDT (dichlorodiphenyltrichloroethane). Based on these values and the $logP_{lit}$ of the standards, a standard curve equation was derived. This equation, in turn, was used to determine the $logP_{TLC}$ for betulin derivatives. The R_{M0} and $logP_{lit}$ [21–23] values for the standard substances A1-A8 are summarized in Table 2.

Table 2. R_{M0} and $logP_{lit}$ values for standard substances.

Standard Substance	Chemical Name	R_{M0}	$logP_{lit}$
A1	acetanilide	0.71	1.21
A2	prednisone	0.85	1.62
A3	benzophenone	2.02	3.18
A4	testosterone	1.72	3.32
A5	anthracene	3.03	4.45
A6	dibenzyl	3.64	4.79
A7	9-phenylanthracene	3.98	6.01
A8	DDT	4.67	6.38

The correlation equation obtained based on data from Table 2; the statistical parameters describing this equation are presented below:

$$logP_{TLC} = 1.2637\ (\pm 0.0883)\ R_{M0} + 0.6129\ (\pm 0.2582)\quad r = 0.986\quad s = 0.344\quad F = 205 \qquad (3)$$

Using the above equation, experimental $logP_{TLC}$ values were determined for all the betulin derivatives tested.

Lipophilicity parameter values can be determined not only experimentally but also using computer programs, which save time. Our intention was to compare prediction programs using different computational algorithms and select the one that generates results that most closely match the experimental values. Therefore, we selected programs that utilize substructures, which are atoms (atomic contribution methods) or groups of atoms (fragmentation methods) [24]. Summarizing the contributions of such substructures, along with the use of correction rules, allows for the determination of logP values. In our calculations, we also utilized programs that quantify logP by examining the entire molecule using various topological indices or molecular properties.

The theoretical values of the lipophilicity parameter (logP) for the studied triterpenoids were obtained from various commercially available online servers.

For example, iLOGP, XLOGP3, WLOGP, MLOGP, SILICOS-IT, and milogP were determined using SwissADME and Molinspiration software Version 2024.01 [25,26], XLOGP2 and ALOGPs values are available on the Virtual Computational Chemistry Laboratory website [27], and VCCLAB, Virtual Computational Chemistry Laboratory [28], and KOWWIN values were obtained using free EPI Suite 4.1 software [29]. All the lipophilicity data are summarized in Table 3.

Table 3. Lipophilicity data for betulin derivatives, obtained both experimentally and theoretically.

Compound	R_{M0}	$logP_{TLC}$	iLOGP	AlogPs	XLOGP2	XLOGP3	WLOGP	MLOGP	SILICOS-IT	KOWWIN	miLogP
1	6.31	8.59	5.48	6.86	9.93	10.05	9.01	6.40	7.98	10.81	8.69
2	6.06	8.27	5.44	6.66	9.16	9.79	8.62	7.56	7.53	10.40	8.55
3	5.92	8.09	4.59	6.23	8.66	8.86	7.98	5.87	7.84	9.53	7.84
4	6.05	8.26	4.76	6.22	9.19	9.48	8.44	6.14	7.43	9.88	8.26
4a	6.87	9.29	5.75	6.7	10.20	10.53	8.70	6.48	9.23	10.61	8.67
4b	6.53	8.86	5.84	7.14	10.18	10.56	9.48	6.82	9.08	11.59	8.78
4c	6.95	9.40	5.27	7.36	10.59	11.04	9.09	6.65	8.50	11.15	9.02
4d	6.99	9.45	5.83	7.21	10.18	10.52	9.40	6.57	8.40	11.38	8.86
4e	7.16	9.66	6.35	7.16	11.01	11.59	9.42	6.99	8.94	11.95	9.18
4f	8.12	10.87	6.27	7.54	12.30	12.30	10.12	7.24	9.57	12.37	9.29
5	5.35	7.37	4.59	6.11	8.42	9.22	8.05	5.96	7.01	9.39	8.03
5a	6.20	8.45	4.94	6.65	9.43	10.27	8.31	6.31	7.81	10.12	8.52
5b	6.65	9.02	5.79	7.07	9.41	10.30	9.09	6.65	8.65	11.10	8.65
5c	6.56	8.90	5.39	7.26	9.82	10.78	8.70	6.48	8.08	10.66	8.91
5d	6.83	9.24	5.40	7.01	9.41	10.26	9.01	6.40	7.98	10.89	8.74
6	5.89	8.06	4.44	5.45	7.92	8.29	7.41	5.60	6.50	8.52	7.14
6a	5.92	8.09	4.80	6.24	8.94	9.34	7.68	5.96	7.29	9.25	7.79
6b	6.90	9.33	5.20	6.79	8.92	9.37	8.46	6.31	8.13	10.23	8.02
BET	4.51	6.31	4.47	5.34	7.81	8.28	7.00	6.00	6.21	8.18	7.16
28-BET	5.33	7.35	4.93	5.53	8.55	8.86	7.57	6.20	6.85	8.77	7.87

In the group of compounds studied, the lowest theoretical logP values were obtained from calculations using the iLOPG 22.1.2.0 program, while the highest were obtained using the KOWWIN v1.68 program (Table 3).

For compounds **4**, **5**, and **6**, which contain a hydroxyl group at the C-28 position, the experimental lipophilicity values ranged from 7.37 to 8.26, which is lower than the results obtained for their 3,28-disubstituted analogs. Comparing the 28-substituted carboxyacylbetulin derivatives **1** and **2** (one carbon atom) and **4d** and **5d** (two carbon atoms) with their precursors **4** and **5**, it can be seen that the introduction of a carbon chain increases lipophilicity, which increases with chain length. The presence of an alkynyl group in the terminal position (compounds **4a** and **5a**) with the same number of carbon atoms (two atoms) in the substituent results in a decrease in lipophilicity. In the group of alkynyl derivatives of the a and c series, changing the position of the triple bond to the central bond results in a higher value of the lipophilicity parameter. For compounds **4c**, **4e**, and **4f** with a central triple bond in the substituent at the C-28, an increase in lipophilicity is observed with increasing carbon atoms in the terminal group (methyl, cyclopropyl, and phenyl). The relationship between experimental lipophilicity and previously determined antiproliferative activity was also analyzed [16]. The lack of a terminal alkynyl substituent at the C28 position results in lower lipophilicity in a given group of carboxyacyl derivatives (compounds **4**, **5**, and **6**) with simultaneous lower biological activity against cancer cell lines such as leukemia (MV4-11), prostate (PC-3), colon (HCT116), and melanoma (Hs249T).

The results of the experimental lipophilicity obtained were subjected to correlation analysis. The relationships between the experimental lipophilicity values (R_{M0} and $logP_{TLC}$) and those determined using available computer algorithms were particularly interesting. Therefore, we began with the correlations between all the lipophilicity values. The correlation matrix is presented in Table 4. They turned out to be characterized by high coefficient

values. The relationship between R_{M0} and $logP_{TLC}$ had the highest value. This is not surprising, as both values are determined experimentally, and one result arises from the other. The lowest correlation coefficient value was found for the experimental lipophilicity values (R_{M0} or $logP_{TLC}$) and the MLOG value, which in both cases was 0.604. The values of the remaining correlations were above 0.8. All the correlation coefficient values were statistically significant, meaning their significance level was greater than or equal to 0.05. This suggests that lipophilicity values calculated using computer algorithms can be applied (after determining appropriate correlation equations) to predict the lipophilicity values of the tested compounds, bypassing the time-consuming laboratory procedure.

Table 4. Correlations matrix for all lipophilicity data obtained.

	R_{M0}	**$logP_{TLC}$**	**iLOGP**	**ALOGPs**	**XLOGP2**	**XLOGP3**	**WLOGP**	**MLOGP**	**SILICOS-IT**	**KOWWIN**	**miLogP**
R_{M0}	1.000										
$logP_{TLC}$	0.999	1.000									
iLOGP	0.819	0.818	1.000								
ALOGPs	0.877	0.877	0.830	1.000							
XLOGP2	0.877	0.877	0.872	0.860	1.000						
XLOGP3	0.868	0.868	0.882	0.910	0.968	1.000					
WLOGP	0.886	0.886	0.900	0.933	0.908	0.914	1.000				
MLOGP	0.604	0.604	0.814	0.704	0.720	0.747	0.755	1.000			
SILICOS-IT	0.889	0.888	0.881	0.888	0.890	0.882	0.901	0.656	1.000		
KOWWIN	0.892	0.892	0.924	0.953	0.919	0.938	0.990	0.770	0.922	1.000	
miLogP	0.814	0.814	0.866	0.936	0.903	0.952	0.935	0.773	0.858	0.948	1.000

A correlation was also considered between experimentally determined lipophilicity values and anticancer activity values for the tested compounds (Table 5). Due to some missing data, compound **4f** was omitted. The values for MiaPaca and Hs2495 were omitted for all betulin-derived compounds. Almost all correlations between experimental lipophilicity values and the analyzed antiproliferative activity values are statistically significant. The exception is the correlation between $logP_{TLC}$ and the value for the A549 cell line. The highest correlation coefficients were obtained for the relationships between R_{M0} or $logP_{TLC}$ and the activity values for the MV4-11 and PC-3 cell lines. These coefficients were −0.737 and −0.728 or 0.727, respectively, for R_{M0} and $logP_{TLC}$. These relatively high correlation coefficients indicate an inversely proportional relationship between the experimental lipophilicity values, expressed using R_{M0} or $logP_{TLC}$, and the antiproliferative activity values for the tested compounds. Based on the data, these relationships can be described by linear equations. In turn, these equations could be used to predict the lipophilicity values of the base of the antiproliferative activity of relevant cell lines.

Table 5. Correlation matrix of experimental values of lipophilicity and anticancer activity for the betulin derivatives analyzed.

	R_{M0}	**$logP_{TLC}$**	**MV4-11**	**A549**	**MCF-7**	**PC-3**	**HCT116**
R_{M0}	1.000						
$logP_{TLC}$	0.999	1.000					
MV4-11	−0.737	−0.737	1.000				
A549	−0.027	−0.028	0.260	1.000			
MCF-7	−0.692	−0.692	0.918	0.448	1.000		
PC-3	−0.728	−0.727	0.942	0.241	0.884	1.000	
HCT116	−0.634	−0.634	0.858	0.522	0.847	0.814	1.000

Other physicochemical parameters describing the tested compounds were also analyzed, including MW (molar mass), TPSA (topological surface area), nROTB (the number of rotatable bonds), nHD (the number of donors of hydrogen bonds), nHA (the number of acceptors of hydrogen bonds), and pharmacokinetic parameters such as logKp (skin permeability), LogPS (central nervous system permeability), and LogBB (blood–brain barrier permeability). The values of these parameters are summarized in Table 6 [30].

Table 6. Values of selected physicochemical and pharmacokinetic parameters of the analyzed betulin derivatives.

Compound	MW [g/mol]	TPSA [Å^2]	nROTB	nHD	nHA	Log Kp	Log PS	LogBB
1	626.91	89.90	8	1	5	−2.735	−2.227	−0.570
2	612.88	89.90	9	1	6	−2.734	−2.176	−0.525
3	584.83	89.90	9	1	6	−2.734	−2.324	−0.521
4	584.87	83.83	7	2	4	−2.735	−1.556	−0.213
4a	639.90	89.90	8	1	5	−2.735	−2.275	−0.554
4b	664.95	89.90	10	1	5	−2.735	−2.271	−0.591
4c	650.93	89.90	8	1	5	−2.735	−2.220	−0.548
4d	640.93	89.90	9	1	5	−2.735	−2.225	−0.605
4e	676.96	89.90	8	1	5	−2.735	−2.165	−0.510
4f	713.00	89.90	8	1	5	−2.735	−2.058	−0.516
5	570.84	83.83	7	2	5	−2.734	−1.513	0.153
5a	622.87	89.90	9	1	6	−2.734	−2.224	−0.508
5b	650.93	89.90	11	1	6	−2.734	−2.169	−0.503
5c	636.90	89.90	9	1	6	−2.734	−1.666	0.106
5d	626.91	89.90	10	1	6	−2.734	−2.173	−0.559
6	542.79	83.83	7	2	5	−2.734	−2.276	−0.538
6a	594.82	89.90	9	1	6	−2.734	−2.320	−0.558
6b	622.87	89.90	11	1	6	−2.734	−2.220	−0.545
BET	442.72	40.46	2	2	2	−2.737	−1.770	−0.444
28-BET	484.75	46.53	3	1	3	−2.614	−2.244	−0.326

The presented parameters include those related to the assessment of the drug similarity of chemical compounds based on the Lipinski rule (molecular weight < 500 Da, lipophilicity < 5, hydrogen bond donors < 5, and hydrogen bond acceptors < 10) and Weber's rule (total polar surface area within 140 Å; number of rotatable bonds < 10) [31]. The compounds tested exceeded both criteria of Lipinski's rule while meeting the requirements of Weber's rule.

The logKp results presented in Table 6, ranging from −2.614 (to −2.737), indicate relatively low skin permeability for all the tested molecules. The in silico blood–brain barrier permeability (LogBB) for the tested triterpenes was <0.3, indicating their inability to cross the BBB. Only four of the tested substances (compounds **4**, **5**, **5c**, and betulin) could penetrate the central nervous system, as their LogPS values were >−2 [30]. Similar results were reported by Stępnik et al. for another triterpenoid, oleanolic acid. Despite the predicted low blood–brain barrier permeability, this compound may affect the central nervous system, reducing brain damage in ischemic stroke, and may also be helpful in neurodegenerative disorders [32].

Table 7 shows the correlation matrix between experimental lipophilicity and ADME parameters. The correlations between R_{M0} and $logP_{TLC}$ and the values of their analyzed properties show that the molar mass of a compound might be the most useful for predicting lipophilicity. In this case, the correlation coefficient is 0.917 and 0.916 for R_{M0} and $logP_{TLC}$, respectively. Among the analyzed properties, except MW, only TPSA and nROTB show statistically significant correlations with the experimentally determined lipophilicity values.

Table 7. Correlation matrix between experimental values of lipophilicity, and the physicochemical and pharmacokinetic parameters for the betulin derivatives analyzed.

	R_{M0}	$logP_{TLC}$	MW	TPSA	nROTB	nHD	nHA	LogKp	LogPS	LogBB
R_{M0}	1.000									
$logP_{TLC}$	0.999	1.000								
MW	0.917	0.916	1.000							
TPSA	0.692	0.692	0.848	1.000						
nROTB	0.636	0.635	0.778	0.897	1.000					
nHD	−0.196	−0.197	−0.081	0.269	0.440	1.000				
nHA	0.422	0.423	0.359	0.0678	−0.162	−0.940	1.000			
LogKp	−0.306	−0.305	−0.462	−0.624	−0.530	−0.241	0.022	1.000		
LogPS	−0.307	−0.308	−0.241	−0.236	−0.329	0.029	−0.084	−0.130	1.000	
LogBB	−0.330	−0.331	−0.232	−0.164	−0.247	0.196	−0.230	0.129	0.846	1.000

Based on the correlation obtained, the best were selected and described using correlation equations. These equations can be applied to calculate lipophilicity values based on theoretically obtained data. The resulting equations and the statistical parameters describing them are in Table 8.

Table 8. Selected correlation equation describing the relationships between R_{M0} or $logP_{TLC}$ and other data.

Dependent Variable	Equation	r	s	F	Number of Equation
R_{M0}	0.8794 (±0.1082) WLOGP − 1.1872 (±0.9320)	0.886	0.374	66	(4)
	0.7684 (±0.0933) SILICOS-IT + 0.7684 (±0.0933)	0.889	0.370	68	(5)
	0.6130 (±0.0733) KOWWIN + 0.0149 (±0.7624)	0.892	0.365	70	(6)
	−0.0474 (±0.0105) MV4-11 + 6.6205 (±0.1352)	−0.737	0.476	20	(7)
	−0.0269 (±0.0062) PC-3 + 6.8227 (±0.1696)	−0.728	0.484	19	(8)
	0.2224 (±0.0637) nROTB + 4.5532 (±0.5343)	0.636	0.632	12	(9)
	0.0114 (±0.0012) MW − 0.5814 (±0.7163)	0.917	0.324	95	(10)
	0.0383 (±0.0094) TPSA + 3.1210 (±0.8063)	0.692	0.583	17	(11)
$logP_{TLC}$	1.1112 (±0.1367) WLOGP − 0.8874 (±1.1774)	0.886	0.472	66	(12)
	0.9703 (±0.1182) SILICOS-IT + 0.9287 (±0.9455)	0.888	0.468	67	(13)
	0.7745 (±0.0926) KOWWIN + 0.6324 (±0.9635)	0.892	0.462	70	(14)
	−0.0957 (±0.0133) MV4-11 + 8.9787 (±0.1710)	−0.737	0.602	20	(15)
	−0.0340 (±0.0078) PC-3 + 9.2341 (±0.2146)	−0.727	0.612	19	(16)
	0.2809 (±0.0804) nROTB + 6.3674 (±0.6754)	0.635	0.789	12	(17)
	0.0144 (±0.0015) MW − 0.1194 (±0.9062)	0.916	0.408	94	(18)
	0.0484 (±0.0119) TPSA + 4.5573 (±1.0190)	0.692	0.737	17	(19)

Based on the obtained results, a similarity analysis was performed for the tested compounds. The first analysis focused on all the lipophilicity values obtained (experimental and theoretical). The corresponding graph is presented in Figure 1. It shows three main clusters: the first composed of compounds **5b**, **5c**, and **5d**; the second composed of compounds **4b**, **4c**, and **4d**; and the third cluster consisting of compounds **3**, **4**, **5**, **5a**, and **6a**, and betulin. Compounds **4f** and 28-acetylbetulin are the most distant from the others. In the case of the latter two, this is due to significant structural differences with respect to the other derivatives. Compound **4f** also has the highest molar mass of all compounds and the highest lipophilicity value in the group of tested compounds. It is also the only one containing an additional substituent in the form of a benzene ring.

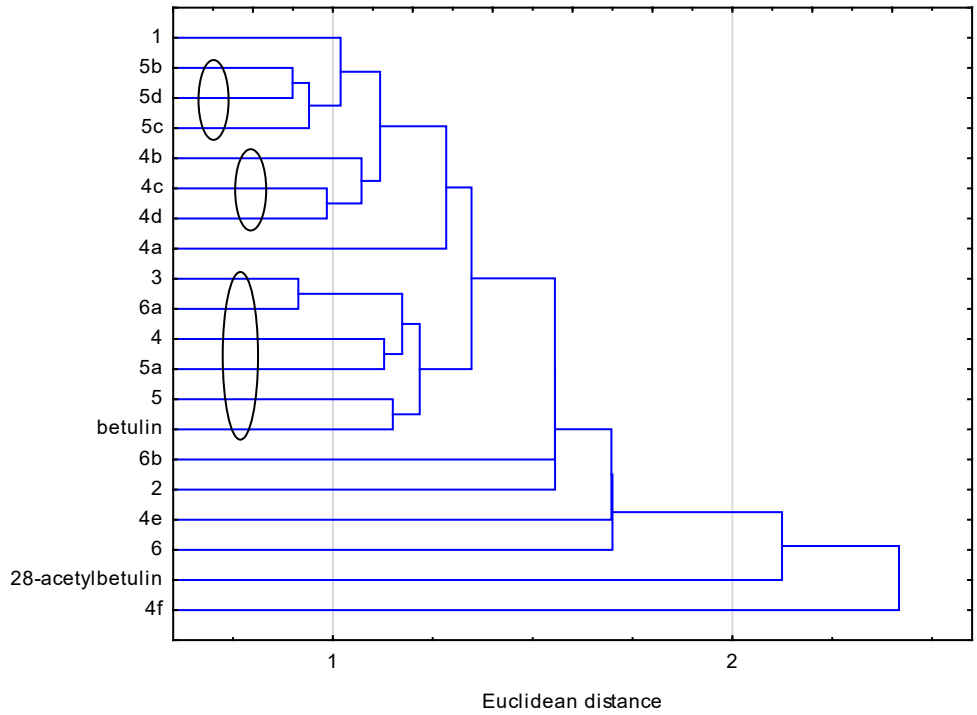


Figure 1. Cluster analysis of lipophilicity values for the betulin derivatives analyzed.

Due to the varying scales of the variables, the standardized data were used for the subsequent CA. To consider the similarities among all the analyzed data describing the compounds, a dendrogram was created. It is presented in the following figure (Figure 2). In this case, the analysis did not include compounds **5a** and **4f** due to missing data. These data describe anticancer activity, i.e., A549, HCT116, Hs249T, and Mia Paca-1 for compound **4f**, and also PC-3, HCT116, and Mia Paca-1 for compound **5a**. Several clusters can be distinguished in the figure: the first is composed of compounds **1** and **4a**–**4d**; the second is composed of compounds **2**, **3**, **6**, **6a**, **6b**, and **5a**–**5d**; the third cluster is composed of compounds **4** and **5**; and the fourth cluster is composed of 28-acetylbetulin and betulin.

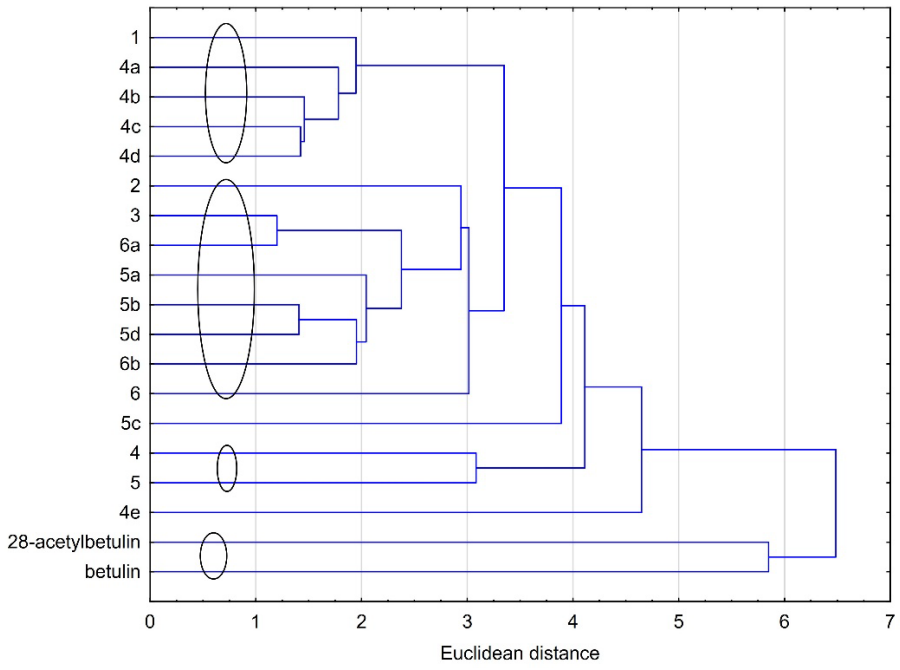


Figure 2. Cluster analysis of all data describing the compounds analyzed.

The next analysis conducted was PCA (principal component analysis), which applied the standardized data as well. This analysis extracted 17 eigenvalues. We evaluated the number of eigenvalues using both the Kaiser criterion and a scree plot (Figure 3), and, finally, the decision was based on the scree plot, which indicated seven eigenvalues. These eigenvalues describe the system's variability as 97%, which is 5% greater than when the Kaiser rule is used. The next table (Table 9) presents the eigenvalue number, the eigenvalue itself, and the percentage of variance for subsequent eigenvalues.

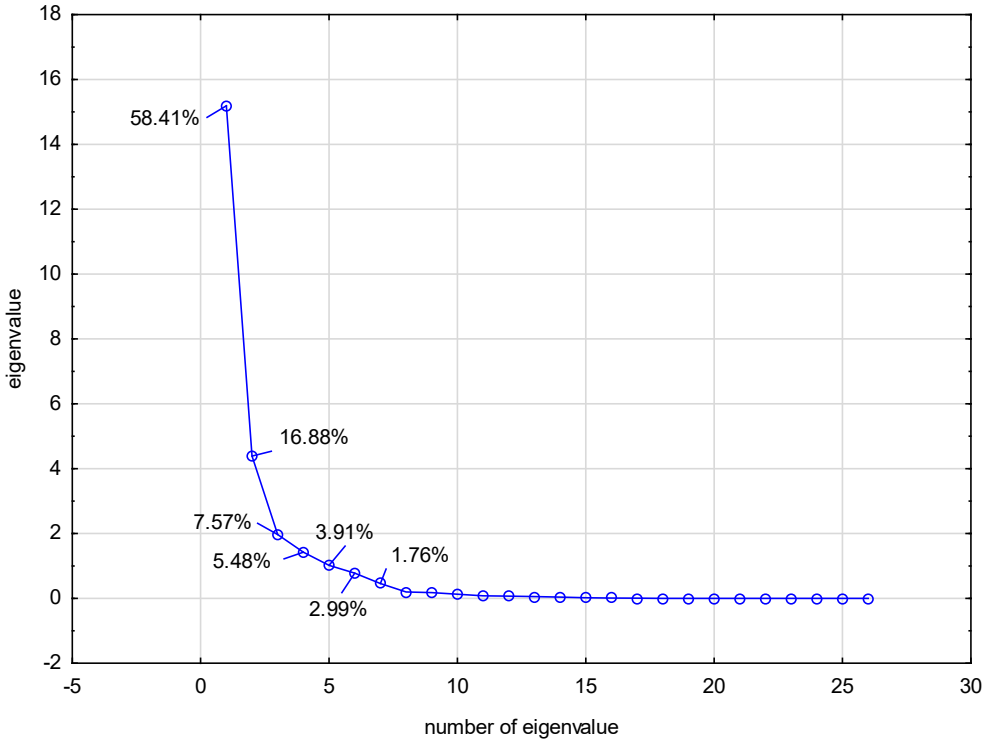


Figure 3. Scree plot for data analyzed.

Table 9. Eigenvalues considered for compounds investigated.

Number of Eigenvalue	Eigenvalue	% of Total Variance
1	15.188	58.41
2	4.388	75.29
3	1.969	82.86
4	1.424	88.34
5	1.017	92.25
6	0.778	95.24
7	0.457	97.00

The projection of the cases onto the factor plane was created based on the obtained results, as shown in Figure 4. The analysis clearly distinguished two groups of compounds: betulin and 28-acetylbetulin, and a second group consisting of all the other compounds analyzed. It indicates that betulin and 28-acetylbetulin differ significantly from the others. This is confirmed by both their structure and the values of all the properties considered. The PCA also confirmed the previous cluster analysis based on the values of all the data for all the compounds (Figure 2). It is evident that betulin and 28-acetylbetulin form a separate cluster, distinct from the other compounds.

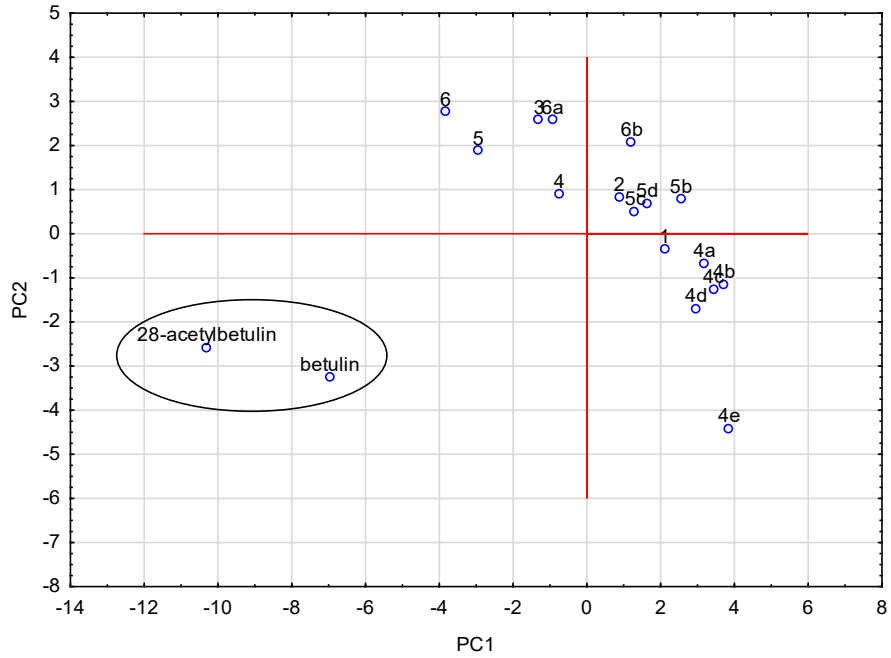


Figure 4. The projection of the cases onto the factor plane for data analyzed.

A nonparametric method, the sum of ranking differences (SRD) analysis was also applied to the studied data. Figure 5 illustrates the conducted analysis. Because low values of the sum of squared differences indicate that the statistical model fits the data well, all values on the left side of the graph, from TPSA to nROTB, properly describe the studied system. Taking into account that this group also includes lipophilicity parameters, both experimental and calculated using computer algorithms, they can be considered as appropriate parameters for describing the lipophilicity of the tested group of betulin derivatives.

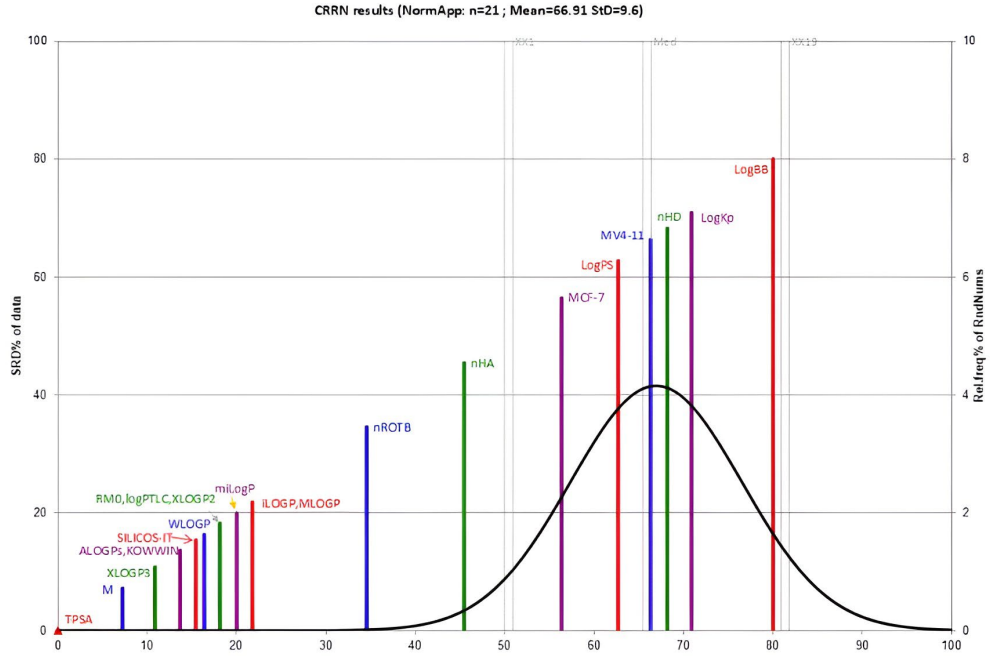


Figure 5. Graphical representation the sum of ranking differences analysis (SRD) for data analyzed.

The SRD analysis performed for the same data corresponds well with the cluster analysis (Figure 6). The values with the smallest sums of squared differences form one cluster, i.e., indicating good similarity among them.

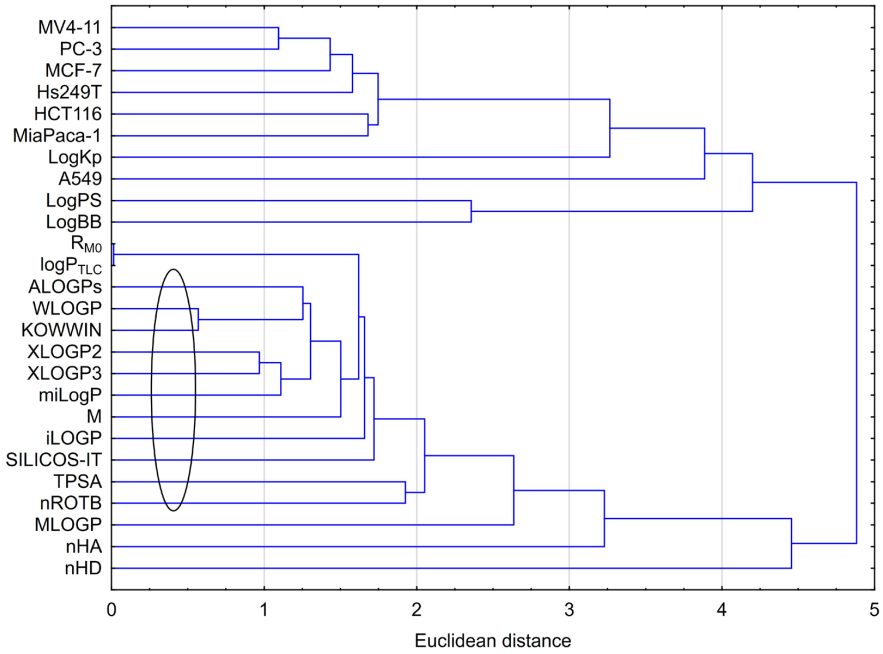


Figure 6. Cluster analysis for all data analysed grouping the parameters describing the betulin derivatives analysed.

The sum of rank differences has previously been successfully used in other studies, for example, to assess the lipophilicity of alkyl and cycloalkyl s-triazine derivatives [33] or antioxidant compounds [34]. Chemometric analysis methods similar to those used in our study (CA, PCA, and SRD) were used for thiepino [3,2-c:6,7-c′]dipyridine and 16-benzothiepino [3,2-c]pyridine derivatives as potential antifungal drugs. For these groups of compounds, lipophilicity was also determined using thin-layer chromatography and in silico methods [35]. As in our case, the compounds studied by the authors can be described by the lipophilicity values determined theoretically.

4. Conclusions

The lipophilicity parameter logP is largely dependent on the structure and number of substituents in the chemical compound molecule. Choosing a calculation method that provides logP values that most closely match the experimental results, while requiring validation on other terpenoid groups, could be helpful in researching these types of compounds.

Lipophilicity values were determined for a group of 3,28-disubstituted betulin derivatives using thin-layer chromatography. It was observed that the type of carboxyacyl substituent at the C-3 position did not significantly affect the lipophilicity of the disubstituted derivatives. The introduction of a substituent (propynoyl group in compounds **4a**, **5a**, and **6a**), which confers significant antiproliferative activity against MV4-11, PC-3, and Hs249T cells, increased lipophilicity compared to the monosubstituted analogs. A decrease in lipophilicity was observed for the alkynyl substituent at the terminal position (compounds **4a** and **5a**) compared to saturated substituents with the same number of carbon atoms (two carbon atoms in compounds **4d** and **5d**). In reality, the relationship between lipophilicity and antiproliferative activity may not be so clear cut. This requires further research, particularly an assessment of the mechanism of the anticancer activity of the compounds studied. This work provides a basis for such studies in the group of biologically active pentacyclic triterpenes of the lupane type. Theoretical lipophilicity values were also determined using available, free software. A range of other data was also obtained for the tested compounds, including physicochemical and pharmacokinetic parameters. The

relationships between the experimental lipophilicity values (R_{M0} and $logP_{TLC}$) and the remaining data describing the analyzed group of compounds were examined.

Basic analysis methods were used to analyze all data: correlation, similarity analysis, principal component analysis, and one of less frequently used method, the sum of ranking differences analysis. The correlation analysis revealed a significant correlation between R_{M0} and $logP_{TLC}$, as well as the theoretically determined lipophilicity values. In the case of anticancer activity values, these correlations were not as high, and some data, i.e., compound **4f** and values of MiaPaca as well as Hs2495, had to be omitted from this analysis due to a lack of proper values. In the case of physicochemical and pharmacokinetic data, only some of them yielded statistically significant correlations. The possibility of using correlation equations to determine lipophilicity values (R_{M0} and $logP_{TLC}$) was assessed using the remaining data values. Mathematical equations were determined (Table 8) for the selected data, namely those with the highest correlation coefficients. The two multivariate analysis methods used, CA and PCA, demonstrated parallel similarity between compounds (Figures 2 and 4). PCA allowed for the reduction of the number of data to seven through the use of eigenvalues. The SRD analysis, on the other hand, allowed for the selection of the data that best described the studied group of compounds. This selection coincides with the analysis of similarities between the data. The cluster created in this case consists of the same data as extracted by SRD analysis. These are: ALOGPs, KOWWIN, WLOGP, XLOGP2, XLOGP3, iLOGP, MLOGP, milogP, SILICOS-IT, M, TPSA, and nROTB. For most of them, the correlations also have high coefficient values, and it was possible to determine correlation equations to calculate the R_{M0} and $logP_{TLC}$ values based on them.

Author Contributions: E.C., K.B.-M. and E.B.: conceptualization, methodology, validation, writing—original draft preparation, project administration, funding acquisition, software, visualization, methodology and writing—review and editing; M.Ś.: writing—review and editing, J.W.: writing—review and editing, supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Medical University of Silesia, grant number (BNW-1-048/K/5/F).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding authors.

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	structure–activity relationship
QSAR	quantitative structure–activity relationship
MW	molecular weight
RP-TLC	reversed-phase thin-layer chromatography
RP-18	octadecylsilane
CA	cluster analysis
PCA	principal component analysis
SRD	sum of ranking differences
DDT	dichlorodiphenyltrichloroethane
TPSA	topological surface area
nROTB	number of rotatable bonds
nHD	number of donors of hydrogen bonds
nHA	number of acceptors of hydrogen bonds
logKp	skin permeability
LogPS	central nervous system permeability
LogBB	blood–brain barrier permeability

References

1. Amiri, S.; Dastghaib, S.; Ahmadi, M.; Mehrbod, P.; Khadem, F.; Behrouj, H.; Aghanoori, M.R.; Machaj, F.; Ghamsari, M.; Rosik, J.; et al. Betulin and its derivatives as novel compounds with different pharmacological effects. *Biotechnol. Adv.* **2020**, *38*, 107409. [CrossRef] [PubMed]
2. Patrick, G.L. Part D Tools of the trade. In *An Introduction to Medicinal Chemistry*, 4th ed.; Oxford University Press: New York, NY, USA, 2009; pp. 377–382.
3. Samiulla, D.S.; Vaidyanathan, V.V.; Arun, P.C.; Balan, G.; Blaze, M.; Bondre, S.; Chandrasekhar, G.; Gadakh, A.; Kumar, R.; Kharvi, G.; et al. Rational selection of structurally diverse natural product scaffolds with favorable ADME properties for drug discovery. *Mol. Divers.* **2005**, *9*, 131–139. [CrossRef] [PubMed]
4. Demets, O.V.; Takibayeva, A.T.; Kassenov, R.Z.; Aliyeva, M.R. Methods of Betulin Extraction from Birch Bark. *Molecules* **2022**, *27*, 3621. [CrossRef] [PubMed]
5. Jalmakhanbetova, R.I.; Mukusheva, G.K.; Abdugalimov, A.S.; Zhumagalieva, Z.Z.; Dehaen, W.; Anthonissen, S.; Suleimen, Y.M.; Seidakhmetova, R.B. Synthesis and Investigation of Biological Activity of New Betulonic Acid Derivatives Containing 1,2,3-Triazole Fragments. *Molecules* **2024**, *29*, 3149. [CrossRef]
6. Ali-Seyed, M.; Jantan, I.; Vijayaraghavan, K.; Bukhari, S.N.A. Betulinic acid: Recent advances in chemical modifications, effective delivery, and molecular mechanisms of a promising anticancer therapy. *Chem. Biol. Drug Des.* **2016**, *87*, 517–536. [CrossRef]
7. Pozharitskaya, O.N.; Karlina, M.V.; Shikov, A.N.; Kosman, V.M.; Makarov, V.G.; Casals, E.; Rosenholm, J.M. Pharmacokinetics and tissue disposition of nanosystem entrapped betulin after endotracheal administration to rats. *Eur. J. Drug Metab. Pharmacokinet.* **2017**, *42*, 327–332. [CrossRef]
8. Jaroszewski, B.; Jelonek, K.; Kasperczyk, J. Drug Delivery Systems of Betulin and Its Derivatives: An Overview. *Biomedicines* **2024**, *12*, 1168. [CrossRef]
9. Sousa, J.L.C.; Freire, C.S.R.; Silvestre, A.J.D.; Silva, A.M.S. Recent Developments in the Functionalization of Betulinic Acid and Its Natural Analogues: A Route to New Bioactive Compounds. *Molecules* **2019**, *24*, 355. [CrossRef]
10. Drag-Zalesińska, M.; Wysocka, T.; Borska, S.; Drąg, M.; Poręba, M.; Choromańska, A.; Kulbacka, J.; Saczko, J. The new esters derivatives of betulin and betulinic acid in epidermoid squamous carcinoma treatment-In vitro studies. *Biomed. Pharmacother.* **2015**, *72*, 91–97. [CrossRef]
11. Cai, S.; Guo, X.; Yang, H.; Zhao, T.; Li, Y.; Deng, N.; Gao, Z.; Meng, O.; Li, X.; Wang, S. Synthesis and antitumor effects of novel betulinic acid derivatives bearing electrophilic moieties. *Bioorg. Med. Chem.* **2025**, *119*, 118062. [CrossRef]
12. Khan, D.A.U. Descriptors and their selection methods in QSAR analysis: Paradigm for drug design. *Drug Discov. Today* **2016**, *21*, 1291–1302. [CrossRef]
13. Andric, F.; Bajusz, D.; Rácz, A.; Segane, S.; Héberger, K. Multivariate assessment of lipophilicity scales—Computational and reversed phase thin-layer chromatographic indices. *J. Pharm. Biomed. Anal.* **2016**, *127*, 81–93. [CrossRef]
14. Šegan, S.; Krunić, M.J.; Andrić, D.B.; Šukalović, V.B.; Penjišević, J.Z.; Jevtić, I.I. Evaluation of lipophilicity and drug-likeness of donepezil-like compounds using reversed-phase thin-layer chromatography. *Biomed. Chromatogr.* **2024**, *38*, e5867. [CrossRef] [PubMed]
15. Soares, J.X.; Santos, Á.; Fernandes, C.; Pinto, M.M.M. Liquid Chromatography on the Different Methods for the Determination of Lipophilicity: An Essential Analytical Tool in Medicinal Chemistry. *Chemosensors* **2022**, *10*, 340. [CrossRef]
16. Chrobak, E.; Świtalska, M.; Wietrzyk, J.; Bębenek, E. New Difunctional Derivatives of Betulin: Preparation, Characterization and Antiproliferative Potential. *Molecules* **2025**, *30*, 611. [CrossRef] [PubMed]
17. Sziklai, B.R.; Héberger, K. Apportionment and districting by Sum of Ranking Differences. *PLoS ONE* **2020**, *15*, e0229209. [CrossRef]
18. Andrić, F.; Héberger, K. Chromatographic and computational assessment of lipophilicity using sum of ranking differences and generalized pair-correlation. *J. Chromatogr. A* **2015**, *1380*, 130–138. [CrossRef]
19. Škrbić, B.; Héberger, K.; Durišić-Mladenović, N. Comparison of multianalyte proficiency test results by sum of ranking differences, principal component analysis, and hierarchical cluster analysis. *Anal. Bioanal. Chem.* **2013**, *405*, 8363–8375. [CrossRef]
20. Héberger, K.; Škrbić, B. Ranking and similarity for quantitative structure–retention relationship models in predicting Lee retention indices of polycyclic aromatic hydrocarbons. *Anal. Chim. Acta* **2012**, *716*, 92–100. [CrossRef]
21. Bodor, N.; Gabanyi, Z.; Wong, C.K. A New Method for the Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786. [CrossRef]
22. Berthod, A. Band broadening inside the chromatographic column: The interest of a liquid stationary phase. *J. Chromatogr. A* **2006**, *1126*, 347–356. [CrossRef] [PubMed]
23. Mannhold, R.; Cruciani, G.; Dross, K.; Rekker, R. Multivariate analysis of experimental and computational descriptors of molecular lipophilicity. *J. Comput. Aided Mol. Des.* **1998**, *12*, 573–581. [CrossRef] [PubMed]
24. Kujawski, J.; Popielarska, H.; Myka, A.; Drabińska, B.; Bernard, M.K. The log P Parameter as a Molecular Descriptor in the Computer-aided Drug Design—An Overview. *Comput. Methods Sci. Technol.* **2012**, *18*, 81–88. [CrossRef]
25. Available online: http://swissadme.ch (accessed on 21 May 2025).

26. Available online: https://www.molinspiration.com/services/logp.html (accessed on 5 January 2024).
27. Tetko, I.V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.A.; Radchenko, E.V.; Zefirov, N.S.; Makarenko, A.S.; et al. Virtual computational chemistry laboratory-design and description. *J. Comput. Aid Mol. Des.* **2005**, *19*, 453–463. [CrossRef]
28. Available online: https://vcclab.org (accessed on 12 September 2025).
29. *Estimation Programs Interface Suite™ for Microsoft® Windows*, v 4.11; United States Environmental Protection Agency: Washington, DC, USA, 2017.
30. Available online: https://biosig.lab.uq.edu.au/pkcsm (accessed on 21 May 2025).
31. Sarkar, P.; Alheety, M.A.; Srivastava, V. Molecular Docking and ADMET Study of Spice-Derived Potential Phytochemicals Against Human DNA Topoisomerase III Alpha. *Macromol. Symp.* **2023**, *407*, 2200108. [CrossRef]
32. Stępnik, K.; Kukula-Koch, W.; Plazinski, W.; Rybicka, M.; Gawel, K. Neuroprotective Properties of Oleanolic Acid—Computational-Driven Molecular Research Combined with In Vitro and In Vivo Experiments. *Pharmaceuticals* **2023**, *16*, 1234. [CrossRef]
33. Salaković, B.; Kovačević, S.; Karadžić-Banjac, M.; Jevrić, L.; Podunavac-Kuzmanović, S.; Antonović, D. Comparative chemometric analysis, ranking and selection of lipophilicity parameters of 6-chloro-1,3,5-triazine derivatives with acyclic and cyclic substituents. *Acta Period. Technol.* **2022**, *53*, 88–99. [CrossRef]
34. Sima, I.A.; Kot-Wasik, A.; Wasik, A.; Namieśnik, J.; Sârbu, C. Assessment of Lipophilicity Indices Derived from Retention Behavior of Antioxidant Compounds in RP-HPLC. *Molecules* **2017**, *22*, 550. [CrossRef]
35. Šegan, S.; Božinović, N.; Opsenica, I.; Andrić, F. Consensus-based comparison of chromatographic and computationally estimated lipophilicity of benzothiepino [3,2-*c*]pyridine derivatives as potential antifungal drugs. *J. Sep. Sci.* **2017**, *40*, 2089–2096. [CrossRef]