Pharmacological classification of anticancer drugs applying chromatographic retention data and chemometric analysis

Marcin Gackowski¹ · Marcin Koba¹ · Robert Pluskota¹ · Emilia Daghir-Wojtkowiak² · Paulina Szatkowska-Wandas¹ · Stefan Kruszewski³

Received: 14 April 2020 / Accepted: 25 July 2020 / Published online: 2 August 2020
© The Author(s) 2020

Abstract

The combination of chromatography and chemometrics has been introduced in order to provide information about drug analytes, biological macromolecules, the stationary phase, whose properties are related to molecular pharmacology and rational drug design. In this work a pharmacological classification was made in order to find strategy for predicting an activity of anticancer drugs. Principal Component Analysis (PCA) method has been employed to build some relationship models between the following: lipophilicity parameters of selected antitumor drugs obtained by chromatographic analysis, some molecular modeling descriptors from HyperChem software, lipophilicity parameters calculated by proper program, biological activity, as well as mechanism of actions and therapeutic properties of selected anticancer drugs. The most significant influence on factors value had parameters, which described compounds' lipophilicity. Important debt on antitumor properties had some other molecular descriptors, which were considered in this paper (e.g. topological, physico-chemical, and energy ones). In addition, distribution of individual drugs on plots determined by two principal components had shown points in good accordance with their chemical structures and cytostatic activity as well as their mechanism of actions and therapeutic properties.

Keywords Principal component analysis (PCA) · Anticancer drugs · Pharmacological classification

Introduction

Chromatography provides qualitative and quantitative information about analyzed sample. Qualitative information involves specified location of the peak in the chromatogram so that it is possible to recognize the type of separated substances, chemical structure of the molecules, physico-chemical properties, and the number of peaks indicating the number of components in the mixture (Malviya et al. 2010).

The data obtained during the chromatographic analysis as well as data from other complex analyses are often assisted by chemometric methods. Chemometric is the field of science and technology based on extraction of useful information from a multi-dimensional data matrix, employing methods from statistics and mathematics. Chemometrics is considered as essential and indispensable tool in many areas of pharmacy such as drug synthesis, development and design, structure–activity relationships, and drug mechanism. Chemometrics is a breakthrough in searching for new drugs, because it incorporates mathematical accuracy, statistical indices, and biological and/or chemical information. With the help of diverse software and chemometric methods, researchers have obtained faster and more accurate results, which represent an exponential progression in Drug Discovery (Kaliszan 1998; Marini 2010; Mocak 2012; Scotti and Scotti 2015). The combination of chromatography and chemometric has been introduced in order to provide information about drug analytes, biological macromolecules, the stationary phase, whose properties are related to molecular pharmacology and rational drug design (Booth and Wainer

✉ Marcin Gackowski
marcin.gackowski@cm.umk.pl

¹ Department of Toxicology and Bromatology, Faculty of Pharmacy, Collegium Medicum of Nicolaus Copernicus University, Bydgoszcz, Poland

² Department of Biopharmacy and Pharmacokinetics, Faculty of Pharmacy, Gdańsk Medical University, Gdańsk, Poland

³ Medical Physics Division, Biophysics Department, Faculty of Pharmacy, Collegium Medicum of Nicolaus Copernicus University, Bydgoszcz, Poland

Fig. 1 a–n The chemical structures of the tested compounds (**a** nimustine, **b** actinomycin D, **c** irinotecan, **d** daunorubicin, **e** doxorubicin, **f** idarubicin, **g** melphalan, **h** mitomycin D, **i** vinorelbine, **j** pirarubicin, **k** docetaxel, **l** vincristine, **m** vinblastine, **n** etoposide)

Fig. 1 (continued)

1996; Norinder and Hermansson 1991; Wolf et al. 1988). Proper processing of chromatographic data can increase the speed and efficiency of searching for new drugs. Whole process is based on the assumption on quantitative relationship between the chemical structure of xenobiotics and their ability for intermolecular interactions. At the same time, it reduces the number of tests involving living organisms. In addition, chromatographic and activity data can reveal information about the drug and investigate the compound's affinity to the stationary phase (Booth and Wainer 1996; Kaliszan 1998; Wolf et al. 1988). Moreover, chemometric methods facilitate processing of the data for further analyses.

Principal Component Analysis (PCA) is the most commonly used chemometric technique. Norinder et al. have applied PCA, to extract the most important factors, further used to establish the regression equation in order to predict the enantioselectivity α , whereas Kumar et al. have applied PCA for classifying aqueous herbal drugs as well as diagnosis and therapeutic prognosis of oral sub-mucous fibrosis (Bober et al. 2011; Koba and Bączek 2012) (Koba and Bączek 2010; Koba and Bączek 2012a, b; Koba et al. 2010b; Norinder and Hermansson 1991; Stasiak et al. 2010; Koba et al. 2010a; Kumar 2017).

Table 1 Comparison of lipophilicity parameter— $\log k_w$ for HPLC columns, used in chromatographic analysis

Compound	$\log k_w$ Luna (A1)	$\log k_w$ Cadenza (A2)	$\log k_w$ TSK-gel (A3)	$\log k_w$ Ascentis (A4)	$\log k_w$ Unison (A5)
Nimustine	0.64	0.81	0.61	0.53	0.95
Actinomycin D	2.01	2.03	2.06	2.03	2.05
Irinotecan	1.23	1.35	1.34	1.21	1.39
Daunorubicin	1.44	1.54	1.57	1.40	1.52
Doxorubicin	1.30	1.41	1.40	1.25	1.38
Idarubicin	1.52	1.62	1.65	1.44	1.61
Melphalan	1.28	1.36	1.29	1.29	1.36
Mitomycin C	0.83	0.88	0.80	0.86	0.92
Vinorelbine	1.43	1.77	1.61	1.35	1.57
Pirarubicin	1.50	1.55	1.55	1.51	1.63
Docetaxel	1.88	1.89	1.9	1.91	1.83
Vincristine	1.29	1.55	1.33	1.21	1.40
Vindesine	1.06	1.20	1.06	1.09	1.26
Vinblastine	1.33	1.79	1.47	1.26	1.45
Etoposide	1.37	1.40	1.36	1.40	1.38

Table 2 Comparison of lipophilicity parameter— $\log k_0$ for HPLC columns, used in chromatographic analysis

Compound	$\log k_0$ Luna (A1')	$\log k_0$ Cadenza (A2')	$\log k_0$ TSK-gel (A3')	$\log k_0$ Ascentis (A4')	$\log k_0$ Unison (A5')
Nimustine	0.54	0.72	0.51	0.45	0.84
Actinomycin D	1.91	1.93	1.95	1.93	1.96
Irinotecan	1.09	1.21	1.17	1.07	1.26
Daunorubicin	1.29	1.41	1.41	1.25	1.41
Doxorubicin	1.15	1.28	1.24	1.11	1.27
Idarubicin	1.36	1.49	1.48	1.29	1.49
Melphalan	1.16	1.25	1.15	1.17	1.26
Mitomycin C	0.74	0.81	0.70	0.76	0.85
Vinorelbine	1.28	1.63	1.43	1.20	1.45
Pirarubicin	1.35	1.42	1.40	1.35	1.51
Docetaxel	1.77	1.78	1.78	1.80	1.74
Vincristine	1.14	1.41	1.17	1.07	1.29
Vindesine	0.93	1.08	0.92	0.95	1.15
Vinblastine	1.18	1.63	1.30	1.12	1.34
Etoposide	1.26	1.30	1.24	1.29	1.29

Chemometric methods are also useful in case of increasing signal-to-noise ratio, removing undesired effects from data, or peak alignment. With the use of projection method and clustering, multidimensional chemical data can be effectively explored. Various strategies of chemometric data are invaluable when calibration and classification models are to be developed (Stasiak et al. 2010). In addition, chemometric methods facilitate the interpretation of the data by identifying relevant variables, e.g. potential biomarkers, organic and inorganic impurities in active substances and ready-to-use drug forms (Cserháti and Szőgyi 2012).

Materials and methods

HPLC analysis

In this study, lipophilicity data for 15 selected anti-cancer compounds, particularly: (1) Nimustine, (2) Actinomycin D, (3) Irinotecan, (4) Daunorubicin, (5) Doxorubicin, (6) Idarubicin, (7) Melphalan, (8) Mitomycin C, (9) Vinorelbine, (10) Pirarubicin, (11) Docetaxel, (12) Vincristine, (13) Vindesine, (14) Vinblastine and (15) Etoposide, were used (Fig. 1a–n). These drugs are derived from many taxonomic groups, and

Table 3 Comparison of lipophilicity parameter—*S*-value, for HPLC columns, used in chromatographic analysis

Compound	<i>S</i> -value Luna (A1'')	<i>S</i> -value Cadenza (A2'')	<i>S</i> -value TSK-gel (A3'')	<i>S</i> -value Ascentis (A4'')	<i>S</i> -value Unison (A5'')
Nimustine	2.00	1.91	2.05	1.56	2.21
Actinomycin D	2.00	1.86	2.14	1.98	1.79
Irinotecan	2.94	2.72	3.43	2.97	2.51
Daunorubicin	2.98	2.57	3.34	2.93	2.26
Doxorubicin	2.9	2.55	3.29	2.83	2.18
Idarubicin	3.05	2.66	3.39	2.94	2.36
Melphalan	2.38	2.22	2.66	2.40	2.05
Mitomycin C	1.84	1.58	2.11	1.91	1.45
Vinorelbine	3.08	2.86	3.53	2.93	2.37
Pirarubicin	3.01	2.44	3.11	3.09	2.44
Docetaxel	2.24	2.05	2.43	2.23	1.88
Vincristine	2.95	2.75	3.15	2.76	2.27
Vindesine	2.65	2.29	2.81	2.80	2.24
Vinblastine	2.98	3.17	3.40	2.83	2.24
Etoposide	2.24	2.02	2.45	2.29	1.77

Table 4 Comparison of log *P* parameters for tested compounds

Compound	AC log <i>P</i> (B1)	miLog <i>P</i> (B2)	KOWWIN (B3)	XLOGP2 (B4)	XLOGP3 (B5)	Average log <i>P</i> (B6)
Nimustine	0.27	0.21	− 0.02	0.34	0.39	0.31
Actinomycin D	− 0.07	0.78	− 0.75	0.95	3.76	1.20
Irinotecan	3.72	4.10	2.33	3.14	2.98	3.30
Daunorubicin	1.43	0.89	2.19	0.14	1.83	1.06
Doxorubicin	0.48	0.57	1.85	− 0.47	1.27	0.51
Idarubicin	1.54	0.89	2.10	0.23	1.86	1.12
Melphalan	1.44	0.08	0.39	0.36	− 0.52	0.39
Mitomycin C	− 1.26	− 0.87	− 1.18	− 1.58	− 0.40	− 0.97
Vinorelbine	3.82	5.92	4.84	3.96	3.59	4.57
Pirarubicin	1.79	2.05	4.58	0.47	2.67	1.79
Docetaxel	2.92	4.24	2.83	2.40	1.65	2.69
Vincristine	3.14	4.95	3.11	2.82	2.82	3.36
Vindesine	2.23	3.09	3.79	2.28	2.55	2.70
Vinblastine	3.38	5.56	4.32	3.87	3.70	4.23
Etoposide	0.53	0.70	0.04	1.04	0.60	0.57

many differences in their mechanisms of action and therapeutic effects were reported. All cytostatics used in experiment were purchased from Sigma-Aldrich (St. Louis, USA) and were used as solutions of a concentration 0.1 mg/ml dissolved in DMSO.

The solvents applied in HPLC analysis were as follows: acetonitrile, water and formic acid, obtained from POCH (Gliwice, Poland). To perform chromatographic measurements, HPLC Shimadzu system (Kyoto, Japan) was used. HPLC was equipped with degasser (DGU-20A5), eluent feed pump (LC-20AD) and autosampler (SIL-20AD) equipped with syringe, detector (SPD-M20A), column oven (CTO-20AC), and recorder (CBM-20A). Data were collected with the use of Solution LC software (version 1.0.0.1 Kyoto, Japan). The compounds were analyzed using the following chromatographic columns: Luna 5u C18 (2) 100 A 150×4.6 mm (Phenomenex, Torrance, CA, USA), Cadenza CD-C18 150×4.6 mm size (Imtakt Corporation, Kyoto, Japan), TSK-gel ODS-80Ts 150×4.6 mm (Tosoh Corporation, Tokyo, Japan), Ascentis C18 150×4.6 mm (Supelco, Bellefonte, PA, USA) and Unison UK-C18 150×4.6 mm (Imtakt Corporation, Kyoto, Japan).

The compounds were analyzed using RP-HPLC system with a gradient elution utilizing solvent A (water + 0.1% formic acid) and solvent B (acetonitrile + 0.1% formic acid). Measurement of the retention times (R_t) was performed with a linear gradient from 5–100% solvent B at two different time points (t_g was 10 or 30 min). Obtained retention times were used as input data to determine the values of $\log k_w$ by DryLab 6.0 program (Monar-Institute, Berlin, Germany). The parameters of $\log k_w$, $\log k_0$, S -value determined for the test drugs analyzed on five different stationary phases are shown in Tables 1, 2 and 3. All measurements were performed at a flow rate of 1 ml/min and wavelength of 254 nm. The volume of eluent contained in the system between the point where the gradient is formed and the point, at

which the mobile phase is introduced into the column, was 0.745 ml. The dead time for all HPLC columns was determined by injecting solvent B. The dead time for Luna C18, Cadenza CD-C18, TSK-gel ODS-80Ts, Ascentis C18 and Unison UK-C18 was estimated at 2.123 min; 1.837 min; 2.621 min; 1.918 min and 1.754 min, respectively.

Lipophilicity parameters

Parameters, which characterized the lipophilicity ($\log P$) of the test compounds were calculated using Virtual Computational Chemistry Application Laboratory (VCCLab, <https://www.vccclab.org/>). To establish the ability for interactive prediction of $\log P$ values, algorithms such as Pharma (parameter AC_log P), Actelion, Molinspiration, accompanying software Dragon X (parameters miLOGP, KOW-WIN) and XLOGP2 and ACOGP3 programs (parameters XLOGP2, XLOGP3) were used (Table 4). The chemical structure of the tested cytostatic was established by molecular modeling using HyperChem 8.0.8 software (Hypercube, Gainesville, USA) and Dragon (Talete, Milan, Italy). Initially, using procedure based on Molecular Mechanics Force Field (MM+), a preliminary optimization of structure of the analyzed compounds was conducted and the results were further analyzed using semi-empirical AM1 method involving the Polak-Ribiere algorithm and the limit of decrease of energy values determined at the 0.01 kcal Å⁻¹.

Structural descriptors

Structural parameters such as total energy (TE), the binding energy (BE), unit atomic energy (IAE), electron energy (EE), core–core interaction energy (CCI energy), heat flow (HF) and the energy of the highest occupied molecular orbital (HOMO) and lowest occupied molecular orbital (LUMO), were

Table 5 Values of descriptors obtained from HyperChem QSAR properties module

Compound	HE (C1)	$\log P$ (C2)	R (C3)	P (C4)	MA (C5)	V (C6)	VdWS (C7)	SAS (C8)
Nimustine	− 13.42	− 0.28	67.19	25.47	272.69	767.27	281.54	482.68
Actinomycin D	− 6.22	− 3.50	326.86	126.99	1255.44	2908.82	1244.18	1322.42
Irinotecan	− 6.68	− 0.46	166.56	61.92	586.69	1570.11	587.23	881.36
Daunorubicin	− 17.88	− 2.40	137.10	51.37	527.53	1309.05	493.23	720.69
Doxorubicin	− 24.44	− 2.96	138.81	52.00	543.53	1308.63	497.41	708.33
Idarubicin	− 17.57	− 1.41	130.73	48.89	497.50	1236.04	467.12	684.63
Melphalan	− 8.13	0.85	81.56	30.56	305.20	863.69	320.26	525.72
Mitomycin C	− 12.40	− 1.42	83.27	31.75	334.33	870.84	324.91	513.38
Vinorelbine	− 3.47	− 1.06	224.57	83.50	778.95	1926.04	755.73	942.93
Pirarubicin	− 20.29	− 1.98	161.59	61.04	627.64	1581.46	597.23	865.92
Docetaxel	− 16.24	1.85	211.68	81.21	807.89	1902.13	761.81	934.58
Vincristine	− 0.94	− 2.36	230.20	86.25	824.97	2003.65	792.69	971.42
Vindesine	− 7.22	− 1.66	216.31	81.48	755.95	1863.90	731.04	914.79
Vinblastine	− 6.17	− 1.46	230.01	86.17	810.99	1991.69	786.62	973.56
Etoposide	− 25.24	− 4.27	145.81	55.15	588.56	1457.28	548.37	805.67

Table 6 Values of descriptors obtained from the HyperChem Software

Compound	TDM (D1)	oX (D2)	oY (D3)	oZ (D4)	Cm + (D5)	Cm - (D6)	HOMO (D7)	LUMO (D8)	TE (D9)	BE (D10)	IAE (D11)	EE (D12)	C-C I (D13)	HF (D14)
Nimustine	4.219	2.369	0.255	3.481	0.4029	-0.3677	1.9262	1.0013	-82,695.3	-3034.8	-79,660.5	-475,936.9	393,241.6	6.6
Actinomycin D	8.942	-4.807	-0.599	-7.517	0.3926	-0.4211	1.9156	0.7477	-385,851.1	-17,863.4	-367,987.7	-6,487,770.7	6,101,919.5	-478.5
Irinotecan	11.806	-5.097	9.898	3.928	0.3954	-0.3807	1.9174	0.8195	-172,907.0	-8569.5	-164,337.5	-1,633,929.8	1,461,022.8	-140.9
Daunorubicin	6.370	-4.068	1.369	4.707	0.3276	-0.3277	1.8795	0.8561	-167,599.8	-7189.3	-160,410.6	-1,538,881.9	1,371,282.0	-355.7
Doxorubicin	5.944	-4.125	2.198	3.673	0.3265	-0.3373	1.8823	0.8233	-174,990.2	-7290.0	-167,700.1	-1,642,372.0	1,467,381.9	-396.9
Idarubicin	5.205	-5.003	1.093	0.933	0.3216	-0.3501	1.9125	0.8996	-156,628.2	-6819.0	-149,809.2	-1,401,052.7	1,244,424.5	-320.0
Melphalan	1.807	0.615	1.697	0.069	0.3120	-0.3655	1.9850	0.9937	-85,684.9	-3653.5	-82,031.4	-522,733.1	437,048.2	-91.0
Mitomycin C	5.180	-2.682	2.359	-3.752	0.3890	-0.4163	1.8699	0.7900	-105,975.7	-4334.4	-101,641.4	-786,001.4	680,025.7	-83.4
Vinorelbine	3.322	2.176	0.413	2.476	0.3070	-0.3732	1.9140	0.7354	-228,132.4	-11,577.3	-216,555.2	-3,056,072.6	2,827,940.2	-145.2
Pirarubicin	2.945	-2.326	-0.912	-1.560	0.3268	-0.3284	1.9127	0.7794	-199,699.4	-8676.6	-191,022.8	-1,974,623.9	1,774,924.5	-452.6
Docetaxel	3.099	0.204	-2.665	1.568	0.4050	-0.3990	1.9174	0.8198	-251,957.2	-11,502.8	-240,454.4	-3,224,419.8	2,972,462.6	-446.3
Vincristine	5.870	0.166	5.453	2.165	0.3127	-0.3600	1.8685	0.7411	-246,534.1	-12,088.0	-234,446.1	-3,343,176.2	3,096,642.1	-261.8
Vindesine	4.219	1.669	3.806	0.729	0.3138	-0.4091	1.9165	0.8271	-220,629.4	-11,480.0	-209,149.4	-2,968,120.3	2,747,490.9	-180.0
Vinblastine	2.343	2.280	0.253	0.481	0.3135	-0.3640	1.9143	0.7723	-239,783.3	-12,101.2	-227,682.1	-3,243,106.0	3,003,322.7	-230.3
Etoposide	4.527	-1.901	1.148	3.944	0.3053	-0.3291	1.8810	0.8042	-191,834.8	-7862.5	-183,972.2	-1,813,111.0	1,621,276.2	-465.2

Fig. 2 Scree plot that shows which factors should be classified to further analysis (based on experimentally determined lipophilicity parameters)

calculated by HyperChem. In addition, value of the highest positive (Cm+) or negative (Cm-) charged electron atoms, distribution of dipole moments with respect to axes X, Y, Z (α_x , α_y , α_z), as well as the total dipole moment (TDM) were calculated. The above-described descriptors were extended by parameters obtained from the “QSAR Properties” module of HyperChem 7.5 Software (HyperCube Inc.), particularly particle surface area available to the solvent (SAS), the surface of Van der Waals bonds (VDWS), molecular volume (V), hydration energy (HE), refraction (R), polarizability (P), atomic weight (MA) and the parameter of lipophilicity (log P). The data are summarized in Tables 5 and 6. Statistical analysis was performed with the use of R® Software (<https://www.r-project.org/>). The input log k_w data have been centered prior to PCA. Presented data have the same units of measure so covariance matrix has been computed from the mean centered data matrix by R® software. All data before the statistical analysis were standardized and the analyses were performed at the significance level of 5% ($\alpha=0.05$).

Results

Statistical analysis

Classification of compounds using principal component analysis (PCA) based on experimentally determined lipophilicity parameters

Principal Component Analysis was performed for data from Tables 1, 2 and 3, with the use of R® Software. To investigate how many factors are statistically significant, auxiliary

scree plot (Fig. 2) was presented. PCA analysis indicated a set of four factors for the interpretation. According to the Kaiser criterion (which assumes that one factor distinct one variable, eigenvalue of the factor should be greater than 1), only two initial factors would be statistically significant. Therefore, the next two should be rejected. The first factor explains 74% of the total variance and the second one 20%. In total, it is more than 94% of all analyzed variables. The second criterion is the scree plot (Fig. 2), for which one should find the point where decrease in variable value is observed. As compared to the Kaiser criterion, scree plot suggests that there are even three statistically significant factors. However, only Factors 1 and 2 were selected for further analysis.

PCA analysis Obtained loadings plot (Fig. 3a) illustrates the relationship between introduced variables and interpreted factors. Analysis of these loadings shows that Factor 1 is more correlated with S-value and its derivatives (which are clustered together), while Factor 2 is correlated with log k_w and log k_0 variables. Highlighted values will be taken into account when interpreting the results.

In Fig. 3b classification of individual groups of drugs is presented. The main criteria were PCA analysis and also the mechanism of action and clinical application of drugs. PCA allowed to group 9 compounds: (10) Pirarubicin, (6) Idarubicin, (4) Daunorubicin, (9) Vinorelbine, (14) Vinblastine, (12) Vincristine, (5) Doxorubicin, (3) Irinotecan. This group contains compounds whose mechanism of action is inhibition of topoisomerase I and II or intercalation in DNA. The resulting grouping contains compounds that are bounded into anthracyclines or alkaloids.

Fig. 3 **a** Two-dimensional scatter plot of the scores presented classification of individual compounds using principal components analysis (PCA) based on experimentally determined lipophilicity parameters. **b** Two-dimensional scatter plot of the scores presented classification of individual compounds using factor analysis (FA) based on experimentally determined lipophilicity parameters

Second generations of anthracyclines (called also anthracycline antibiotics, derived from *Streptomyces* bacterium) were classified into one cluster due to the uniform mode of action, together with alkaloids derived from the periwinkle plant *Catharanthus roseus*—Vinblastine, Vinorelbine and Vincristine. All these mentioned cytostatics are used to treat leukemias, lymphoproliferative units, testicular cancer, breast cancer, lung cancer and bladder cancer. Presented main cluster contained also Irinotecan, which is a semisynthetic analogue of the natural alkaloid camptothecin. Its mechanism of action depends on preventing DNA from relaxing by inhibition of topoisomerase *I*. It is also used in the treatment of breast and lung tumors.

All these clustered variables presented on score plot (Fig. 3b) were characterized by low values of both—Factor 1 and Factor 2.

Classification of compounds based on experimentally determined lipophilicity parameters, together with lipophilicity obtained by computational methods

Principal Component Analysis for the data in Tables 1, 2, 3 and 4, was performed in R[®] Software. To investigate how many factors are statistically significant, the scree plot was presented (Fig. 4). Five factors should have been subjected for further analysis. To extract the most statistically significant factors Kaiser Criterion was applied. There are three factors for which the eigenvalue is near to 1. Although factor 3 explained only 7.7% of the total data variance, however, after summing Factor 1 and Factor 2, more than 86% of the information contained in the input data was included; therefore the analysis will be conducted for Factor 1 and Factor 2.

Fig. 4 Scree plot that shows which Factors should be classified to further analysis (based on experimentally determined lipophilicity parameters, together with lipophilicity obtained by computational methods)

PCA analysis After analysis of loadings (Fig. 5a), it was observed that higher correlation occurs between Factor 1 and *S*-value variables. New interpretable variables are XLOGP2 and Average log *P*, which are also correlated with Factor 1. Moreover, occurrence of such many derivatives of partition coefficient parameter may indicate that Factor 1 depends on increasing lipophilicity and thus increases the solubility of the drug. Factor 2 included mainly log k_w parameter for all used HPLC columns (which may possibly draw conclusion that used columns are sufficiently selective to separate all analytes).

Factor 1 is associated with parameters that allow predicting lipophilicity and its presence will have a decisive influence on the permeability of the drug into the cerebrospinal fluid. The classification will be difficult; however, the main criterion, due to PCA analysis, the mechanism of action and clinical application, remains unchanged.

The proposed method of classification is presented on Fig. 5b. Cluster “A” consists of (15) Etoposide, (7) Melphalan and (5) Doxorubicine, assigned according to their clinical use, main cancer of the hematopoietic tissues. All classified drugs indicated different mechanism of action. Cluster B consists of (4) Daunorubicin, (6) Idarubicin, (10) Pirarubicin, and (11) Docetaxel. Drugs classified into cluster B have similar medical use. This compound is used in therapy of head and neck cancer or breast cancer. Cluster C includes (3) Irinotecan, (12) Vincristine, (14) Vinblastine, (9) Vinorelbine and (13) Vindesine. Cluster C drugs are used in the treatment of lung cancer.

Classification based on all lipophilicity parameters in combination with molecular descriptors

In this case, Principal Component Analysis was performed in R® software for data from Tables 1, 2, 3 and 6. Obtained

eigenvalues and the scree plot (Fig. 6) of the test data set were presented in this section. Lipophilicity parameters in combination with the data obtained by HyperChem 7.5 Software (HyperCube Inc.) were analyzed. After this analysis, six statistically significant factors were isolated. However, after application of the Kaiser criterion, the number of factors was reduced to main two factors, and the total value of variance explained by these factors was about 70%.

PCA analysis As was observed previously, Factor 1 was related to lipophilicity parameters such as log k_w , log k_0 , also, Factor 2 was correlated with *S*-value. The proposed method of classification is presented on Fig. 7a. Hydration energy (HE) was strongly correlated. Moreover, it is responsible for exceeding the energy of the crystal structure and therefore facilitates the dissolution of the compound. Log *P* and hydration energy influenced the ability to dissolve substances and their penetration through biological membranes. Parameters such as refractive index (*R*), polarizability (*P*), atomic weight (mass amu), volume (*V*), surface bonding and Van der Waals surface, were dependent on Factor 1, and these values described (1) the ability of a molecule to deformation under the influence of the electric field, (2) susceptibility to change the angle of rotation, (3) amount of Van der Waals bonds and (4) solvent accessible surface area. This surface is relatively large; thus we could infer that dissolution of the compound should be very simple. Total dipole moment (TDM), which is the vector of the sum of bonds, described compound’s polarity. Parameter, which described Atom’s negative charge depended on Factor 1. The highest HOMO orbital, correlated with Factor 2, was totally filled up by electrons. It plays a significant role in the occurrence of new and breaking old chemical bonds. The total energy (TE), binding energy (BE), electron energy

Fig. 5 **a** Two-dimensional scatter plot of the scores presented classification of individual compounds using principal component analysis (PCA) based on experimentally determined lipophilicity parameters, together with lipophilicity obtained by computational methods; Factor 2 relative to Factor 1. **b** Two-dimensional scatter plot of the scores presented classification of individual compounds using Factor Analysis (FA) based on experimentally determined lipophilicity parameters, together with lipophilicity obtained by computational methods; Factor 2 in relative to Factor 1

(EE) and nuclear energy (IAE) values are positive, so every bond in the molecule is stable and strong, and also depends on value of Factor 1. Additionally, C–C interactions are negative, which can indicate that it could be easy to tear weak bonds between cores. Total energy, binding energy, electric energy or C–C interactions will depend on Factor 1. For Actinomycin D, Vinorelbine, Pirarubicine, Docetaxel, Vincristine and Vinblastine energies are negative, which means that these substances are present in bound form. Hydration energy, correlated with Factor 2, in Daunorubicin, Doxorubicin, Idarubicin, Melphalan, Pirarubicin, Docetaxel and Etoposide, so it is sufficient to disrupt the bonds in the crystal structure and increase solubility defined by total dipole moment, which divides the drugs as less and more polar. The division is dependent on increasing electronegativity between elements, which form the compound.

PCA allowed to observe two groups of compounds (Fig. 7b: Cluster “A” [(4) Daunorubicin, (5) Doxorubicin, (6) Idarubicin, (15) Etoposide, (10) Pirarubicin] and Cluster “B” [(3) Irinotecan, (9) Vinorelbine, (12) Vincristine, (13) Vindesine, (14) Vinblastine]. All drugs from cluster A are used in treatment of acute leukemia and drugs from cluster “B”—in lymphoma.

Discussion

In this study, antitumor activity of 15 test drugs was evaluated. Classification of these cytostatic drugs according to the mechanism of action and clinical application by principal components analysis was conducted. The principle of PCA method is to reduce the large amount of data and present

Fig. 6 Scree plot that shows which Factors should be classified to further analysis (based on experimentally determined lipophilicity parameters in combination with molecular descriptors)

it in the form of several main factors, explaining the total variance. For this purpose, data on retention parameters, lipophilicity parameters and structural descriptors (describing the chemical structure) were calculated. In next step, chemical structure, anti-tumor activity and other properties were discussed, based on the obtained results from PCA.

Figures 3a and 5a showed lipophilicity parameters such as $\log k_w$, $\log k_0$ and S -value. After analysis of both, we assumed that $\log k_w$ and $\log k_0$ values have been correlated with Factor 1. This may indicate that this factor is responsible for drug dissolution as well as better absorption into the tissues. By contrast, Factor 2 is related to S -value. This parameter is responsible for solubility in aqueous or oily solution.

Afterwards, combination of retention parameters together with the values characterizing the lipophilicity $\log P$ have been considered, which led to obtain various results. Lack of the rotation indicated that Factor 1 is connected with XLOGP3 and Average $\log P$ variables. After application of the Varimax rotation, parameters such as AClogP, miLogP, XLOGP2 and Average $\log P$ were assigned to Factor 3. We might assume that the division of these values between two factors may result from antitumor activity. Its strength is primarily based on the impact of Factor 3 and less from other factors. The obtained activity of tested substances strictly depends on the lipophilicity and is likely to be related to the distribution and accumulation in the cells. In pharmacy, $\log P$ is a very important parameter, which can predict rate of drug's distribution in the human body, the dose and toxicity as well.

Considering the juxtaposition of several retention parameters with molecular descriptors, it can be observed that without and with Varimax rotation approach, five statistically significant factors were obtained. Descriptors which described the features of the investigated molecules are replaced by numerical values. In this paper descriptors belonging to the class of constitutional (calculated on the basis of the molecular formula), geometry (describing the size and shape of the molecule), quantum—mechanical (characterized by semi-empirical methods; molecular orbital energies) and electrostatic (which show charge distribution in the molecule) group were used. Geometric configuration of drug's chemical structure may have an important role affecting the properties of the tested anticancer drugs.

In each example, classification of these pharmaceuticals based on their values of charges was proposed. One can postulate that clusters formed allow better rationalization of therapy and contribute to more effective treatment.

Conclusions

The present study confirms that lipophilicity is one of the key parameters characterizing the usefulness of active substances that confirms state of the art in the field of drug design. Moreover, the study indicates that the use of PCA can effectively allow the classification of chemical compounds on the basis of computational data (including lipophilicity) not only due to the spatial structure, but also due to the mechanism of action and potential therapeutic

A

PC2

PC1

B

PC2

PC1

A

B

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bober L, Koba M, Judycka-Proma U, Baczek T (2011) Pharmacological classification of drugs by principal component analysis applying molecular modeling descriptors and HPLC retention data. *J Chromatogr Sci* 49(10):758–763
- Booth TD, Wainer IW (1996) Investigation of the enantioselective separations of α -alkylarylcarboxylic acids on an amylose tris(3,5-dimethylphenylcarbamate) chiral stationary phase using quantitative structure–enantioselective retention relationships identification of a conformation. *J Chromatogr A* 737(2):157–169. <https://www.sciencedirect.com/science/article/pii/S0021967396000118>
- Cserhádi T, Szógyi M (2012) Application of multivariate regression models in chromatography. *Eur Chem Bull* 1(7):274–279
- Kaliszan R (1998) Retention data from affinity high-performance liquid chromatography in view of chemometrics. *J Chromatogr B Biomed Appl* 715(1):229–244
- Koba M, Baczek T (2010) Chemometric evaluation of the significance of molecular structural descriptors on binding of acridinone derivatives to DNA. *Lett Drug Des Discov*. 7(7):494–499. <https://www.eurekaselect.com/openurl/content.php?genre=article&issn=1570-1808&volume=7&issue=7&page=494>
- Koba M, Bączek T (2012) Importance of some classes of molecular descriptors on classification of antitumor acridinones using factor analysis. *Med Chem Res* 21(10):2854–2862
- Koba M, Bączek T (2012) The influence of lipophilicity on the classification of antitumor acridinones evaluated by principal component analysis. *Curr Pharm Anal* 8(2):157–74. <https://www.scopus.com/inward/record.url?eid=2-s2.0-84861778574&partnerID=40&md5=6c3806769adfdde09a3d63cdf249ce57>
- Koba M, Baczek T, Macur K, Bober L, Frackowiak T, Bucinski A et al (2010a) Factor analysis of microbiological activity data and structural parameters of antibacterial quinolones. *J. Mol. Model.* 16(2):327–335
- Koba M, Bober L, Judycka-Proma U, Baczek T (2010b) Influence of HPLC retention data and molecular modeling descriptors on prediction of pharmacological classification of drugs using principal component analysis method. *Comb. Chem. High Throughput Screen* 13(9):765–776
- Kumar K (2017) Principal component analysis: most favourite tool in chemometric. *Resonance* 1:747–759
- Malviya R, Bansal V, Prakash Pal O, Kumar SP (2010) High performance liquid chromatography: a short review. *J Glob Pharma Technol* 9(2):22–26
- Marini F (2010) Classification methods in chemometrics. *Curr Anal Chem* 1(6):72–79
- Mocak J (2012) Chemometrics in medicine and pharmacy. *Nov Biotechnol Chim* 11(1):11–26
- Norinder U, Hermansson J (1991) Chiral separation of N-aminoalkylsuccinimides on an α 1-acid glycoprotein column: a quantitative structure—enantioselectivity relationship study. *Chirality* 3(5):422–426. <https://doi.org/10.1002/chir.530030508>
- Scotti M, Scotti L (2015) Editorial (thematic issue: chemometrics in drug discovery). *Comb Chem High Throughput Screen* 18(8):702–703
- Stasiak J, Koba M, Bober L, Baczek T (2010) Principal component analysis of HPLC retention data and molecular modeling structural parameters of cardiovascular system drugs in view of their pharmacological activity. *Int. J. Mol. Sci.* 11(7):2681–2698
- Wolf RM, Francotte E, Lohman D (1988) Quantitative correlation between calculated molecular properties and retention of a series of structurally related racemates on cellulose triacetate. *J Chem Soc* 6:893–901

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