Cur.Chem.Res.

ISSN : 2278-0963

WWW.MEHTAPRESS.COM

Full Paper

Current
CHEMICAL RESEARCH

Received: 15/07/2011
Accepted: 17/02/2012

Novel anticancer drugs utilizing data mining for substituent substitution within uramustine scaffold

Ronald Bartzatt
University of Nebraska, Durham Science Center, Chemistry Department, 6001 Dodge Street, Omaha, Nebraska 68182, (USA)
E-mail: rbartzatt@mail.unomaha.edu

Abstract

Uramustine (or uracil mustard), is an alkylating chemotherapy drug used to treat lymphatic malignancies such as non-Hodgkin's lymphoma. Ten nitrogen mustard agents that are analogous to uramustine are presented following data mining algorithmic search for substituent/substructure moieties. While ten percent of the outcomes were acceptable for consideration, nevertheless the structures show remarkable analogy to uramustine and strong potential for clinical cancer application. Vital structure features such as nitrogen mustard group and pyrimidine-dione construction are present in ten analogues to uramustine. Important pharmacological properties such as Log P, polar surface area, formula weight, and molecular volume are determined. The range of Log P and polar surface area for the ten analogues are -1.93 to 1.88 and 36.47 A^2 to 121.14 A^2, respectively. All ten analogues to uramustine showed zero violations of the Rule of 5, indicating favorable bioavailability characteristics. Analysis of similarity (ANOSIM) determined that these ten analogues with uramustine are significantly distinct by pharmaceutical properties ascertained. Uramustine was also distinct from all analogues based on hierarchical cluster analysis. Polar surface area is highly correlated (Pearson $r > 0.8200$) to formula weight, number of oxygens & nitrogens, and number of hydroxyls & amines. High resolution discriminant analysis indicated that four analogues were close to the parent compound uramustine in terms of molecular properties. The morbidity and mortality of lymphomas is such a state to strongly advocate the continuous investigation of new anticancer drugs.

Keywords

Uramustine; Uracil mustard; Non-Hodgkin's lymphoma; Nitrogen mustard.

Corresponding author's name and address

Ronald Bartzatt
University of Nebraska, Durham Science Center, Chemistry Department, 6001 Dodge Street, Omaha, Nebraska 68182, (USA)

INTRODUCTION

Uramustine (or uracil mustard, 5-aminouracil mustard, chlorethaminacil), is an bi-functional alkylating nitrogen mustard agent that utilizes a natural product to carry the nitrogen mustard group to a selective metabolite site[1]. The potential for applying uramustine as a therapeutic drug appeared in the 60's[2], where its potential as a treatment tool for metastatic Wilms's tumor was notes[3] (Wilm's tumor or nephroblastoma, is a rare kidney cancer primarily affecting children).

Uramustine has been shown to be effective in the treatment of lymphoma, chronic lymphatic leukemia, and thrombocythemia[4]. In addition, uramustine has been shown to control thrombocytisis without any report to be leukemogenic[5,6]. Uramustine has been shown to have levels of efficacy for the treatment of lymphoreticular tumors[7], chronic granulocytic leukemia[8], lymphosarcoma, and soft tissue sarcoma[9]. Various administration and clinical properties were determined in previous studies[10,6].

Various alkylating agents have been implicated as the cause of pulmonary diseases. One incidence with chronic myelogenous leukemia presented with typical cytology,

biopsy, and roentgenologic findings of lung toxicity after six weeks of therapy with busulfan is noted[11]. These similar roentgenologic changes have also occurred after administration of uracil mustard[11]. This report reiterates the necessity for developing novel drug designs but having the selective site advantage of uracil construction. A previous study showed that a hybrid of uramustine and minor groove binder Distamycin A improved the alkylating properties[12].

These findings and others strongly support the investigation and development of novel anticancer designs having pyrimidine-dione structures. This work demonstrates the efficacy of data mining search approach for limited substituent and substructure replacement. The complex uracil structure can be a useful lead to produce potential antineoplastic pharmaceutical agents.

EXPERIMENTAL

Molecular modeling and assembly of constructs

Molecular modeling (2-D) was accomplished utilizing ACD/ChemSketch modeling v. 10.00 (Advanced Chemistry Development, 110 Yonge Street, Toronto Ontario, M5C 1T4 Canada). In silico structure search for substituent replacement was accomplished using chemical substructure and similarity search with Molsoft L.L.C., and some molecular properties provided by source algorithm (Molsoft L.L.C. 11199 Sorrento Valley Road, S209 San Diego CA 92121). Various properties; polar surface area, violations of Rule of 5, molecular volume, number of oxygens/nitrogens/amines/hydroxyls, Rule of 5, were determined using Molinspiration Properties Calculations module (Molinspiration Chemiformatics, Nova ulica 61, SK-900 26 Slovensky Grob, Slovak Republic). Where appropriate to visualize steric structure characteristics then SPARTAN Modeling was utilized (Wavefunction, 18401 Von Karman Ave., Ste. 370, Irvine CA 92612 USA).

Pattern recognition and elucidation

To identify underlying associations/patterns within the descriptors multivariate numerical data matrix then various pattern recognition techniques were implemented. Included in this analytical approach is hierarchical cluster analysis accomplished by KyPlot v. 2.0 Beta 15 (copyright Koichi Yoshioka 1997-2001). Other pattern recognition elucidation by discriminant analysis and ANOSIM (analysis of similarity) were performed by PAST v. 1.80 (copyright Oyvind Hammer, D.A.T. Harper 1999-2008).

Numerical analysis of multivariate data matrix

Statistical analysis and Pearson r determination of all numerical data was performed by Microsoft EXCEL (EXCEL 2003, copyright 1985-2003). Screening for numerical outliers was done by Grubb's Test (extreme studentized deviate) by GraphPad Software (2236 Avenida de la Playa, La Jolla, CA 92037 USA). Multiple regression analysis was performed by GraphPad InStat v. 3.0 for Windows 95 (HJ Motulsky, GraphPad InStat 3.0 GraphPad Software, Inc., San Diego California USA, www.graphpad.com)

RESULTS AND DISCUSSION

Uramustine is known to be effective in the treatment of lymphoma, chronic lymphocytic leukemia (CLL), and thrombocythemia[4]. Lymphoma is a cancer in the lymphatic cells of the immune system. Usually, lymphomas are present as a solid tumor of lymphoid cells[13]. This type of malignant cells often originate in lymph nodes and as an enlargement of the node (this being a tumor). It can also affect other organs in which case it is referred to as extranodal lymphoma. Extranodal sites include the skin, brain, bowels and bone. Lymphomas are closely related to lymphoid leukemias[13]. There are many types of lymphomas, with lymphomas part of the broad group of diseases called hematological neoplasms[13]. Chronic lymphocytic leukemia induces a slow increase in white blood cells referred to as B lymphocytes (or B cells). These cancerous cells can spread from the blood marrow to the blood, also affecting the lymph nodes or organs such as the liver and spleen. CLL can eventually induce failure of the bone marrow. Thrombocythemia, a myeloproliferative blood disorder, is characterized by the excessive production of platelets in the bone marrow.

Data mining for substructure replacement has undergone progress in recent years and presents an efficacious tool for drug design particularly with the presence of effective parent constructs that are applied in 'seeding' the mining process. Applying uramustine as the parent drug design (see drug 1, Figure 1) the outcome nitrogen mustard agents 2 through 11 were determined from 100 candidates, an outcome of approximately 1 out of 10 success rate (or 10% success). The algorithmic search process generated about 90% failed candidates having no nitrogen mustard group and/or complex scaffolding with unfavorable bioavailability attributes (ie. high formula weight and polar surface area). This rate of ruination advocates the necessity of having information descriptive of successful candidates concerning the physiological health disorder under investigation.

The full molecular structures of uramustine and ten analogues are shown in Figure 1. In general, the pyrimidine-

dione blueprint remains in addition to the nitrogen mustard group that renders the bi-functional alkylating antineoplastic action. Variations include a carboxyl group (-C(O)OH) added to analog 6, alkylated carbamic acid moiety (R-NH- C(O)OH) of analog 5, and the tetrahydro-pyran substituent of analog 11. For analog 8 the nitrogen mustard group (-$N(CH_2CH_2Cl)_2$) forms part of urea substituent (-NC(O)N-).

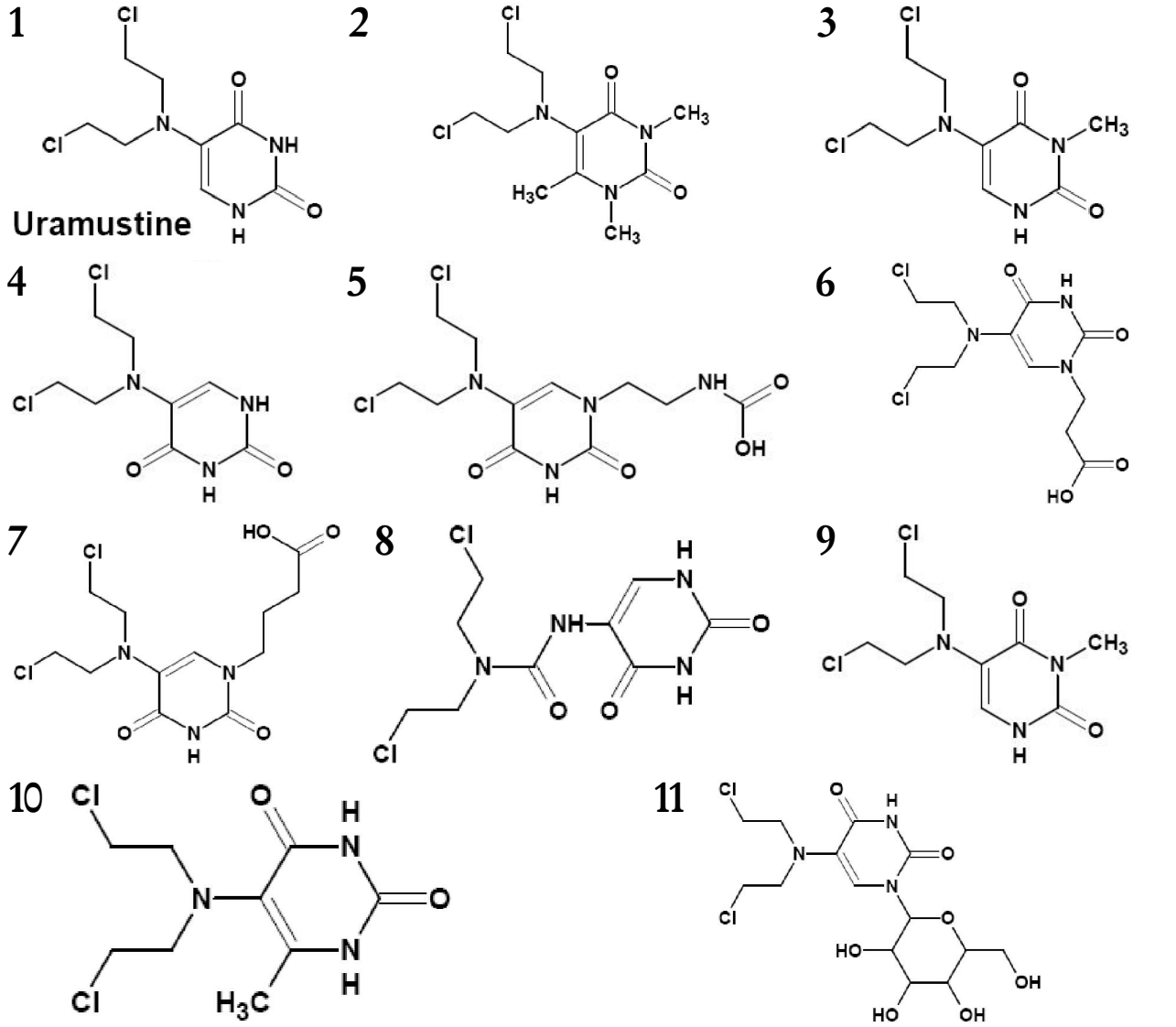

Figure 1 : Comparison of uramustine (1) to molecular structures of analogs (2) through (11) demonstrates the preservation of the nitrogen mustard moiety (-$N(CH_2CH_2Cl)_2$) as well as pyrimidine-dione scaffolding. Other notable variations include a carboxyl group (-C(O)OH) added to analog (6), alkylated carbamic acid moiety (R-NH-C(O)OH) of analog (5), and the tetrahydro-pyran substituent of analog (11).

Numerical quantities for molecular properties, shown in TABLE 1, clearly indicate the diverse range in Log P, polar surface area, etc, that result from substituent substitution. As example, the range in Log P, polar surface area, and formula weight are as follows; -1.930 to 1.880, 36.470 A^2 to 121.14 A^2, and 224.05 to 449.05, respectively. Although the numerical values of these molecular properties are shown to affect bioavailability, this outcome clearly indicates a diverse group of anticancer nitrogen mustard agents that are anticipated to present various effectiveness towards antineoplastic activity (an objective in novel drug design). Positive values in Log P are more lipohilic agents whereas negative values more hydrophilic (ie. More water soluble). Uramustine, determined to have Log P equal to 0.538, is near the mean value of 0.522 for all other analogs. Most hydrophilic are analogs 5 (Log P = -1.93), analog 8 (Log P = -0.03), and analog 11 (Log P = -0.68). Interestingly, the polar surface area (PSA) of uramustine at 68.962 A^2 is at the mean value of 68.01 A^2 for all analogs. The properties Log P and PSA are indicative of druglikeness (PSA reflective of number of oxygens, nitrogens, hydroxyls, and amines) suggested by the Rule of 5[14]. The Rule of 5 describes properties suitable for favorable bioavailability as follows: 1) Not more than 5 hydrogen

bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms); 2) Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms); 3) A formula weight not greater than 500; 4) An octanol-water partition coefficient Log P not greater than 5. All analogs of uramustine, and uramustine, show zero violations of the Rule of 5, hence possessing favorable bioavailability descriptive of druglikeness. Polar surface area is highly correlated (Pearson r > 0.8200) to formula weight, number of oxygens & nitrogens, and number of hydroxyls & amines. Interestingly, Log P values are significantly inversely correlated (Pearson r $\leq$ -0.3800) to PSA, number of atoms, formula weight, number of oxygens/nitrogens/ hydroxyls/amines, and molecular volume.

TABLE 1 : Properties of anticancer drugs

Compound	Log P	Polar Surface Area (A^2)	Number of Atoms	Formula Weight	Number Oxygens & Nitrogens	Number of Hydroxyls & Amines	Volume (A^3)
1, Uramustine	0.538	68.962	13	224.047	5	2	165.44
2	1.88	36.47	35	293.07	5	0	305.2
3	1.49	45.36	29	265.04	5	1	260.09
4	0.9	54.16	26	251.02	5	2	236.89
5	-1.93	95.9	37	374.03	8	3	311.3
6	0.50	75.06	35	323.04	7	2	303.1
7	0.42	74.98	37	337.06	7	2	323.76
8	-0.03	76.3	30	294.03	7	3	284.12
9	1.49	45.36	28	265.04	5	1	260.09
10	1.17	54.38	29	265.4	5	2	259.63
11	-0.68	121.14	47	449.05	10	5	364.77

A^2 = Angstroms^2 A^3 = Angstroms^3

Hierarchical cluster analysis is utilized to analyze large multivariate data sets and arrange subjects (drugs is this study) into groups (clusters) in which members have the highest similarity based on numerical values[15]. Results of cluster analysis are shown by 2-way dendrogram in Figure 2, utilizing single linkage (the distance between two clusters is computed as the distance between the two closest elements in the two clusters) and standard Euclidean distance (distance between two points that would be measured with a ruler). The cluster at origin (node A) separates into subclusters and distinguishes analog 11 and 5 from uramustine and all analogs. Joined at node B, but made separate is the parent uramustine (distinct) from analogs, and subclusters dividing the remaining anticancer agents. Analog 2 is made distinct from subclusters at node C and D having analogs 3, 9, 4, 10; and 6, 7, 8; respectively. Consequently to the outcome suggests that uramustine (drug 1) is distinct from the analogs, and analogs 5, 11, and 2 have a measured level of distinctness. Analogs 3, 9, 4, and 10 and determined to be most similar to each other, suggesting some level of likeness in pharmaceutical activity. Likewise, analogs 6, 7, and 8 are most alike based on molecular properties.

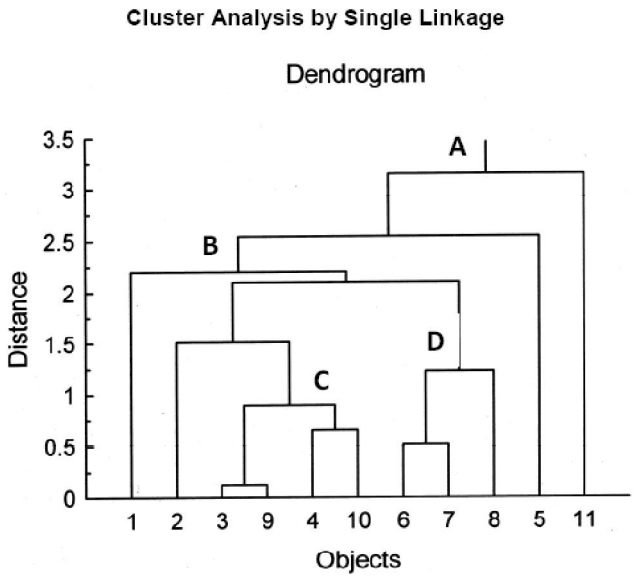

Figure 2 : Hierarchical cluster analysis of properties presented in TABLE 1 show that the parent structure uramustine (1) and analog (11) are distinct from the remaining agents. The analogs (6), (7), and (8) are grouped together, with two miniclusters having (3) and (9) with (4) and (10) in a group.

Analysis of similarity (ANOSIM) is testing for difference between two or more multivariate groups, based on any distance measure, where a R value of zero suggests complete randomness in the data set[16]. However a large positive *R* (up to 1) signifies dissimilarity between groups (drugs in this study), with an outcome of R = 1.00 occurring using the properties of TABLE 1; indicating these drugs are dissimilar.

These results of data mining search for substituents did produce interest trends in the number of oxygens, nitrogens, -OH, and –NHn, which is presented in Figure 3. A linear relationship with formula weight is obtained

for oxygens and nitrogens ($y = 0.0248x - 1.2472$ ($R^2 = 0.8994$)) with formula weight the independent variable. However a polynomial relationship exists between formula weight and quantity of hydroxyls and amines (see Figure 3), that are described with equation $y = 0.0001x^2 - 0.0631x + 10.178$ ($R^2 = 0.6673$). Likewise a polynomial relationship of polar surface area to formula weight is generated by this data mining (see Figure 4; $y = 0.0015x^2 - 0.6836x + 132.05$ ($R^2 = 0.7607$).

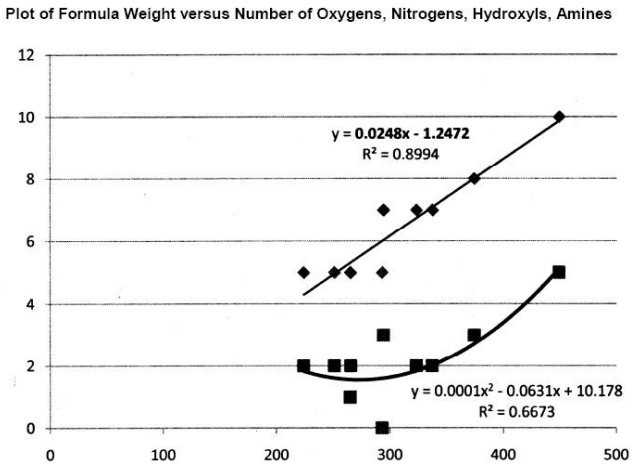

Figure 3 : The two-way plot of formula weight (x-axis) versus number of oxygens, nitrogens, hydroxyls, and amines are shown here and obtained from TABLE 1. The number of oxygens and nitrogens per analog is defined by linear equation $y = 0.0248x - 1.2472$ ($R2 = 0.8994$) as formula weight the independent variable. However the relationship of formula weight to quantity of –OH and –NHn is polynomial $y = 0.0001x2 - 0.0631x + 10.178$ ($R2 = 0.6673$).

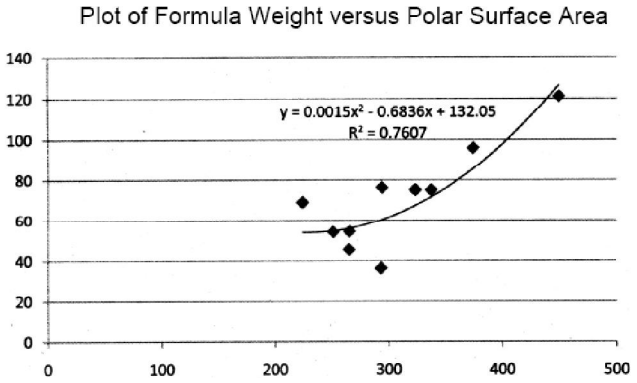

Figure 4 : For this group of anticancer agents the resulting relationship of polar surface area to formula weight is polynomial $y = 0.0015x^2 - 0.6836x + 132.05$ ($R^2 = 0.7607$).

The purpose of a discriminant analysis is to predict group membership and ascertain understanding of the data set (ie. Molecular properties of drugs)[15]. It can identify features which characterize or separate two or more classes of objects or events[15]. Applying discriminant analysis to TABLE 1, divided these agents into two groups wherein each group includes members having highest similarity and greatest distinction from the other group. This outcome formed two groups with 72.73% correct classification: Group 1) uramustine (1), analogs 2, 5, 9, and 11; Group 2) analog 3, 4, 6, 7, 8, and 10. Although a differing outcome than cluster analysis, this outcome can by interpreted to for predictive purposes in describing the importance of the independent variables (descriptors) in differentiating conclusions of known group membership (ie. These descriptors enable the distinction of this group of antineoplastic agents).

To draw a mathematical relationship among uramustine and analogs, somewhat reflective of the process of substituent data mining and the relationship between several independent or predictor variables and a dependent or criterion variable, multiple regression is performed with TABLE 1 descriptors. Designating formula weight (FW), numbers of oxygens & nitrogens (nON) and –OH & -NH_n (nOHNH), with molecular volume (volume), the outcome is: FW = 104.00 + (0.1512)(Log P) + (1.526)(PSA) + (8.154)(number of atoms) + (9.586)(nON) – (15.485)(nOHNH) – (0.6740)(volume). Having R^2 of 0.9928 the model explains 99.28% of the variance in formula weight, with number of atoms as the highest significance for formula weight.

Substituent replacement is an important approach in rational drug design[1]. This approach has proved highly successful for development of powerful antiviral agents based upon acyclovir[17], and very useful antibacterial agents in the class of cephalosporins[18]. Applied toward uramustine, ten antineoplastic agents were developed that retain the bi-functional nitrogen mustard moiety as well as the pyrimidine-dione fundamental structure enabling selective uptake at the site of the tumor. All analogs showed favorable bioavailability based on the Rule of 5. The use of pattern recognition analysis revealed underlying relationships of these drugs based upon the multivariate set of descriptors. Consequently, this set of novel drugs resulted from substituent data mining, with properties suitable for pharmaceutical application towards cancer treatment. Further study and wider application, this approach should reveal novel drug designs for clinical application.

CONCLUSIONS

Ten analogs to uramustine were identified utilizing substituent data mining for nitrogen mustard constructs. The pyrimidine-dione fundamental form remains in addition to the nitrogen mustard group that renders the bi-functional alkylating antineoplastic action. All analogs of uramustine, and uramustine itself, showed zero violations of the Rule of 5, and possessing favorable bioavailability descriptive of druglikeness. ANOSIM

analysis indicated the entire set of drugs are distinct from each other with cluster analysis revealing that uramustine is distinct from all analogs. Analogs 3, 9, 4, 10 were clustered and therefore determined to be most similar; whereas 6, 7, 8 were likewise determined to be most similar to each other. Multiple regression analysis drew a mathematical relationship among the independent variable (descriptors) for formula weight (dependent variable) that is useful for predicting molecular properties of similar agents. Novel drug designs can be procured by this approach that can be useful for the treatment of lymphomas.

ACKNOWLEDGEMENTS

This study is supported by the College of Arts and Sciences, University of Nebraska, Durham Science Center, Omaha, Nebraska 68182 U.S.A.; S.D.G.

REFERENCES

[1] T.Nogrady; 'Medicinal Chemistry A Biochemical Approach', 2nd Edition, Oxford University Press; Oxford, (**1988**).
[2] G.Ballerini, G.Castoldi, N.Ricci, L.Tenze; J.Clin.Ter., **15(32)**, 49 (**1965**).
[3] T.J.Vietti, D.H.Berry, D.J.Fernbach, J.Lusher, W.W.Sutow; Am J.Dis.Child., **108**, 530 (**1964**).
[4] B.J.Dennedy, J.L.Torkelson, E.Torlakovic; Cancer, **85(10)**, 2265 (**1999**).
[5] B.T.Toh, S.A.Gregory, W.H.Knospe; Am.J.Hematol., **28(1)**, 58 (**1988**).
[6] H.K.Shamasunder, S.A.Gregory, W.H.Knospe; JAMA, **244(13)**, 1454 (**1980**).
[7] I.E.Fortuny, A.Theologides, B.J.Kennedy; Minn.Med., **55(8)**, 715 (**1972**).
[8] R.P.Herrmann, L.Dougan, J.A.Holt, J.M.Jackson, M.L.Matthews, A.J.Nelson, N.S.Stenhouse, H.J.Woodliff; Med.J.Aust., **1(16)**, 789 (**1972**).
[9] D.H.Berry, W.W.Sutow, T.J.Vietti, D.J.Fernbach, M.P.Sullivan, M.E.Haggard, D.M.Lane; J.Clin.Pharmacol.New Drugs, **2(4)**, 169 (**1972**).
[10] J.H.Roberson; Blood, **35(3)**, 288 (**1970**).
[11] D.G.Hankins, S.Sanders, F.M.MacDonald, C.W.Drage; Chest, **73(3)**, 415 (**1976**).
[12] P.G.Baraldi, R.Romagnoli, A.E.Guadix, M.J.Pineda de las Infantas, M.A.Gallo, A.Expinosa, A.Martinez, J.P.Bingham, J.A.Hartley; J.Med.Chem., **45(17)**, 3630 (**2002**).
[13] P.Parham; 'The Immune System', Garland Science; New York, (**2005**).
[14] C.A.Linpinski, F.Lombardo, B.W.Dominy, P.J.Feeney; Adv. Drug Del.Rev., **46**, 3 (**2001**).
[15] S.T.Bow; 'Pattern Recognition', Marcel Dekker; New York, (**1984**).
[16] K.R.Clarke; Australian Journal of Ecology, **18**, 117 (**1993**).
[17] B.Golankiewicz, T.Ostrowski, G.Andrei, R.Snoeck, E.De Clercq; J.Med.Chem., **37**, 3187 (**1994**).
[18] K.E.Price, D.N.McGregor; Scand.J.Infect.Dis.Suppl., **42**, 50 (**1984**).