British Journal of Pharmaceutical Research

16(4): 1-12, 2017; Article no.BJPR.33205
ISSN: 2231-2919, NLM ID: 101631759

Pyrimethamine Based Anti-protozoan Agents from Isostere and Heuristic Structure-similarity Search

Ronald Bartzatt[1*]

[1]Durham Science Center, Department of Chemistry, University of Nebraska, 6001 Dodge Street, Omaha, Nebraska 68182, USA.

Author's contribution

Author RB designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript.

Article Information

DOI: 10.9734/BJPR/2017/33205
Editor(s):
(1) Horacio Perez-Sanchez, Department of Computer Engineering, University of Murcia, Spain.
(2) Ali Nokhodchi, Professor of Pharmaceutics and Drug Delivery, School of Life Sciences, University of Sussex, UK.
Reviewers:
(1) Suhas Vasaikar, Baylor College of Medicine, USA.
(2) Dragos Horvath, Université de Strasbourg, France.
(3) Abdullahi M. Nuhu, CST. Kaduna Polytechnic, Kaduna, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/19021

Original Research Article

Received 4th April 2017
Accepted 8th May 2017
Published 11th May 2017

ABSTRACT

Aims: To generate new medicaments for control and treatment of the parasitic protozoan *Toxoplasma gondii.*
Study Design: Structure similarity search and isostere search was conducted over a broad range of structure categories. Correlation and highest similarity scores were implemented to select the best drug candidates.
Place and Duration of Study: University of Nebraska, Department of Chemistry, Durham Science Center, 6001 Dodge Street, Omaha Nebraska 68182, from June 2016 to February 2017.
Methodology: Utilizing pyrimethamine as the parent compound, a broad range of similar structures and isosteres were found by applying search methods. The compounds having the highest correlation and similarity scores were selected for the study of molecular properties. The molecular properties were determined and examined for underlying relationships by pattern recognition hierarchical cluster analysis and K-means cluster analysis.
Results: Thirty compounds were identified to have a very high level of structure similarity or isosteric relationship to pyrimethamine. The molecular structures and molecular properties are

**Corresponding author: E-mail: rbartzatt@unomaha.edu;*

presented for all compounds, inclusive of pyrimethamine. Hierarchical cluster analysis and K-means cluster analysis indicated compounds with highest underlying similarity to pyrimethamine. Box plots showed the over-all distribution of important pharmaceutical properties, such as molecular weight, Log P, polar surface area, number of rotatable bonds, molecular volume, and number of hydrogen bond donors. Structure components are compared to elucidate potential clinical activity. Multiple regression is applied on all compounds to generate a numerical relationship for prediction of similar compounds. Save for only one isostere, all compounds showed zero violations of the Rule of 5, indicating favorable drug-likeness and bioavailability.

Conclusion: Thirty compounds highly analogous to pyrimethamine were identified following heuristic search course. The molecular properties were determined for all compounds and indicated genuine potential for treatment of toxoplasmosis. Correlation of structure and pattern recognition methods indicated 30 compounds of clinical potential and property analogy to pyrimethamine.

Keywords: Protozoan; Toxoplasma gondii; pyrimethamine; antimalarial; isosteres.

ABBREVIATIONS

PSA: Polar surface area; MW: Molecular weight; nOHNH: Number of hydroxyl and amine groups; nON: Number of oxygen and nitrogen atoms.

1. INTRODUCTION

Toxoplasma gondii (*T. gondii*) causes toxoplasmosis, and is an intracellular, obligate, parasitic protozoan [1]. *T. gondii* is found worldwide and is one of the most common parasites in developed countries, with up to 23.6% in places in Africa [1,2,3]. Healthy adults may not show any symptoms of infection, however, individuals having weakened immune systems the infection can cause serious to fatal consequences [2]. Epidemiology studies have shown that up to 84% of individuals in France are infected [4].

Infection with *T. gondii* has been shown to be associated with various neurological disorders, and in particular schizophrenia [5]. Infection with *T. gondii* can occur by eating poorly cooked or raw meat that contains the *T. gondii* tissue cysts [6]. For North America, the greatest threat for infection comes from raw or poorly cooked pork [6]. Other modes of infection include blood transfusions, organ transplant, unpasteurized goat milk, soil contact, and consumption of contaminated vegetables/fruits [6]. The Protozoa group are diverse unicellular eukaryotic organizms, most of which range in size from 10 micrometers to 52 micrometers [7]. These organisms are abundant in aqueous environments and soil [7]. Other protozoan diseases associated with humans include [7]: Malaria, Amoebiasis, Giardiases, Trichomoniasis, Chagas disease, Leishmaniasis, Sleeping Sickness, and Amoebic dysentery.

Antiprotozoal agents (ATC code: ATC PO1) are grouped together as medicinal applied in the clinical treatment of protozoan infection [8]. This group of medicinal agents are distinguish by their mechanism of action or by organism type, they include the following [8,9,10]: eflornithine, furazolidone, metronidazole, ornidazole, pentamidine, pyrimethamine, and tinidazole. Pyrimethamine (5-(4-chlorophenyl)-6-ethyl-2,4-pyrimidinediamine) is well absorbed, administered orally, has a high protein binding capacity, and is excreted by the kidneys [11]. There is widespread resistance to pyrimethamine, and the side effects include nausea, vomiting, headache, and diarrhea [7]. Despite this, pyrimethamine is considered among the most important medications for basic health maintenance. This due to a large part to the serious socio-economic and health burden of *Toxoplasma gondii* proliferation in developed and developing nations [12].

Structure-based drug design is most efficacious when it is a part of a complete drug lead discovery process (i.e. new designs for additional clinical medicaments). Substitution of substituents of a parent compound (simplest member of a group of compounds) pursues the objective of chemical similarity that includes the generation of isosteres. The entity having chemical similarity (or molecular similarity), refers to the similarity of chemical elements with respect to either structural and/or the biological effect of the chemical compound [13].

Computer-aided drug discovery/design methods have played a major role in the development of therapeutically important small molecules [13]. Computer-aided drug design has enabled the focus of drug discovery to achieving the best potency against a biological target of interest, and favorable absorption, distribution, metabolism, and excretion (ADME) [13]. A heuristic approach employs a practical method (not necessarily optimal), but sufficient for the immediate goals. When finding an optimal solution is impractical, heuristic methods can be used to speed up the process of finding a satisfactory solution [13].

Clearly, the study and development of additional drug agents to combat *Toxoplasma* type infections is of great importance. This study presents novel derivatives of pyrimethamine by heuristic structure search and isostere identification. The drug pyrimethamine possesses substituents on the aromatic ring that facilitates and makes efficacious the isosteric substituent of atoms. The compounds that result are shown to have efficacious molecular properties when compared to the parent drug pyrimethamine and by their similarity score.

2. METHODOLOGY

2.1 Properties and Molecular Modeling

For all 31 compounds included in this study, the numerical values of molecular properties were determined by utilizing heuristic techniques of Molinspiration (Molinspiration Cheminformatics, Nova ulica 61, SK-900 26 Slovensky Grob, Slovak Republic). Similarity search and isostere elucidation was accomplished by Molinspiration. Similarity scoring and correlation to pyrimethamine for all structures was accomplished by Molinspiration. Construction and presentation of molecular structure was accomplished by use of ACD ChemSketch Modeling v. 12.01 (Advanced Chemistry Development, 110 Yonge Street, Toronto Ontario, M5C 1T4 Canada) and Molinspiration (http://www.molinspiration.com/services/search.html)

2.2 Pattern Recognition and Multivariate Statistical Analysis

Identification of underlying associations and patterns within the numerical values of molecular properties done by use hierarchical cluster analysis and K-means cluster analysis. Hierarchical cluster analysis was performed by KyPlot v. 2.0 (copyright Koichi Yosioka 1997 to 2001 and K-means cluster analysis was performed by PAST v. 2.06 (copyright Oyvind Hammer, D.A.T. Harper 1999-2008).

2.3 Various Statistical Analysis Data

Methods of statistical analysis applied included Pearson r correlation and univariate descriptive statistics were accomplished by Microsoft EXCEL v. 14.0.6112.5000 (EXCEL Professional plus 2010). Pearson r correlation was also determined by using PAST v. 2.06. Multiple regression by Graph Pad Instat version 3.00 (GraphPad Software, Inc., San Diego, California USA; www.graphpad.com). Box plots were generated by PAST v. 2.06. Grubb's test for outliers accomplished by Graph Pad Quick Calcs (http://www.graphpad.com/ quickcalcs/index.cfm)

3. RESULTS AND DISCUSSION

3.1 Structure and Properties

Therefore, the computer-aided heuristic similarity search accompanied by isostere identification has revealed 30 new molecular structures for the treatment of protozoan *T. gondii* infections. The chemical structure of pyrimethamine with the six isostere compounds is presented for comparison in Fig. 1. The pyrimethamine molecule and six isosteres have various attributes in common. All seven compounds have an aromatic ring and a pyrimidine ring having a primary amine group (or pyrimidin-2-amine group), with a second primary amine group and one ethyl substituent (or 6-ethylpyrimidine-2, 4-diamine group).

The pyrimethamine structure has a 4-chlorophenyl group covalently attached to the 6-ethylpyrimidine-2,4-diamine framework. The isosteres are formed upon substitution of the 4-chloro atom with analogous atoms (i.e. same number of atoms and/or the same number of valence electrons). The selection of the isostere substitutions were selected to optimize drug likeness of the new formed compounds and maintain the attributes of isostere (i.e. maintain same number of atoms/valence electrons). Isosteres are known to be effective for enhancing drug likeness [13,14,15,16].

This accomplished, compound (2) has a phosphane group ($-PH_2$) substituting for the chlorine atom of pyrimethamine drug (see Fig. 1). Following in order, compounds 3, 4, 5, 6,

and 7 have the following atoms substituting for the chlorine atom of pyrimethamine, respectively:

(3) –SH atoms; (4) –F atom; (5) –OH atoms; (6) $–NH_2$ atoms; and (7) $–CH_3$. The alterations these substitutions bring about on the molecular properties are presented in Table 1 and will be discussed next.

For most drugs, the preferred route of administration is by oral ingestion [13]. Therefore, investigators seek the physicochemical properties that favor intestinal absorption [13,14]. The rule-of-5 (RO5) is a rapid screen for compounds likely to fit the parameters of drug-likeness and have favorable membrane permeation and bioavailability [14]. The RO5 was formulated after examination of over 2000 known drugs and observing that orally administered drugs are relatively small and moderately lipophilic molecules [14]. The RO5 states that if a compound violates any two of the following rules, it is likely to exhibit poor intestinal absorption [14]: 1) Molecular weight greater than 500; 2) Number of hydrogen bond donors more than 5 (where, a donor being any O-H or N-H group); 3) Number of hydrogen bond acceptors more than 10 (where, an acceptor being any O or N including those in donor groups); and 5) Calculated Log P greater than 5.0.

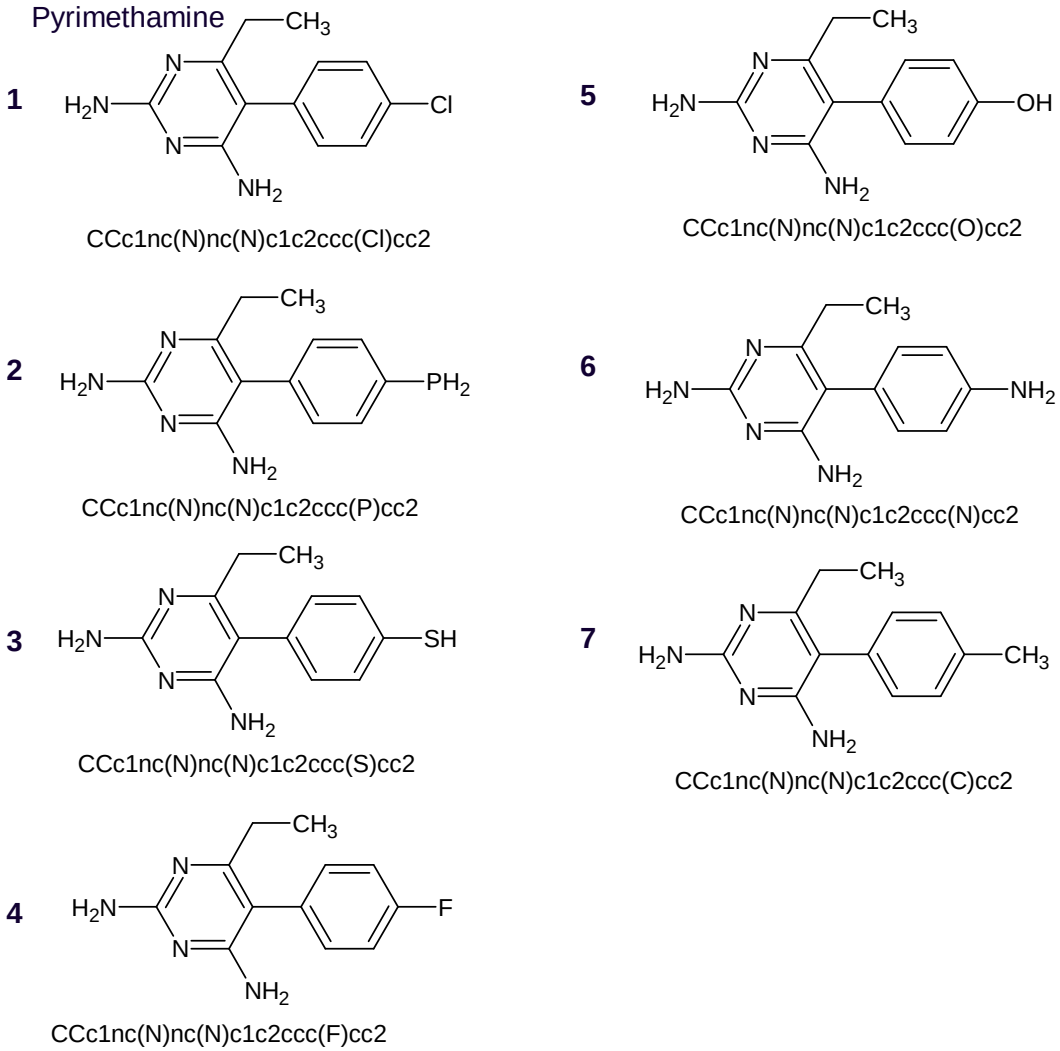


Fig. 1. Isosteres (molecules or ions with the same number of atoms and/or the same number of valence electrons) of pyrimethemine (1)

Table 1. Molecular properties of compounds

Drug	Log P	Polar surface area (Angstroms^2)	Number of atoms	Molecular weight	Number of oxygen & nitrogen atoms	Number of -OH & -NH_n	Violations of rule of 5	Number of rotatable bonds	Volume (Angstroms^3)
1 Pyrimethamine	2.84	77.83	17	248.72	4	4	0	2	216.62
2 isostere	2.82	77.83	17	246.25	4	4	0	2	222.49
3 isostere	2.39	77.83	17	246.34	4	4	0	2	220.74
4 isostere	2.32	77.83	17	232.26	4	4	0	2	208.01
5 isostere	1.68	98.06	17	230.27	5	5	0	2	211.10
6 isostere	1.23	103.85	17	226.29	5	6	1	2	214.37
7 isostere	2.61	77.83	17	228.30	4	4	0	2	216.64
8	2.16	77.83	16	214.27	4	4	0	2	203.09
9	2.85	45.75	17	228.29	3	1	0	3	221.99
10	4.05	17.31	17	242.71	2	0	0	1	210.75
11	3.33	30.20	17	264.12	3	0	0	1	203.57
12	1.48	84.93	17	230.27	5	3	0	2	211.91
13	1.81	54.87	17	250.69	4	1	0	2	210.67
14	3.00	34.38	17	222.25	3	0	0	2	199.64
15	2.51	34.90	18	307.19	3	0	0	3	240.26
16	2.38	34.90	18	262.74	3	0	0	3	235.91
17	1.15	55.12	18	244.29	4	1	0	3	230.39
18	4.05	17.31	17	242.71	2	0	0	1	210.75
19	4.08	34.38	19	270.72	3	0	0	2	229.74
20	2.02	46.09	18	241.29	4	0	0	4	228.83
21	4.28	17.31	17	263.13	2	0	0	1	207.73
22	1.71	45.75	15	200.24	3	1	0	1	188.39
23	1.34	93.62	16	213.24	5	5	0	0	187.84
24	3.77	17.31	17	246.68	2	0	0	1	199.12
25	3.65	17.31	16	228.68	2	0	0	1	194.19
26	3.40	34.38	18	236.27	3	0	0	2	216.20
27	3.45	83.64	18	264.67	6	1	0	3	212.97
28	3.51	17.83	17	242.71	2	0	0	1	210.75
29	1.42	55.12	19	258.32	4	1	0	4	247.19
30	2.55	39.43	18	239.28	4	0	0	2	218.60
31	4.18	17.31	17	287.16	2	0	0	1	215.10

Shown in Table 1, for compounds 1 to 7 in Fig. 1, all seven compounds show favorable bioavailability and membrane penetration (favorable drug-likeness) with zero violations of RO5, with agent 6 showing only one violation of the RO5. This is a beneficial and favorable outcome for the efficacy of these six isosteres of pyrimethamine. In addition, previous studies have shown that drugs having polar surface area (PSA) greater than 140 Angstroms^2 are poorly absorbed [13,15]. Note, that all agents 1 to 7 have PSA less than 103.85 Angstroms^2 and are expected to be well-absorbed following ingestion. Other properties of pyrimethamine have a very strong negative relationship (Pearson's r correlation less than -0.7000) for PSA, number of –OH, -NH_n, oxygen atoms, and nitrogen atoms. The molecular weight (MW) of pyrimethamine is a very strong positive correlation to MW of all six isosteres (Pearson r = 0.7031).

The compounds presented in Fig. 2 are derived from heuristic search for chemical similarity to pyrimethamine. Compounds 8 through 19 showed a similarity score to pyrimethamine of 0.66 to 0.87. The greater the similarity score, the higher the correlation to pyrimethamine. Compounds 8 through 19 have a very broad variation in molecular scaffolding, however, the molecular properties are still showing zero violations of RO5, indicating favorable drug-likeness to bioavailability.

Compounds 8 to 19 show aromatic rings, hetero-atomic aromatic rings, halogen atoms, and aliphatic substituents. Overall the variation in structure is broad, however consistency in molecular properties with the features of pyrimethamine is clear and paramount.

The compounds presented in Fig. 3 are derived from heuristic search for chemical similarity to pyrimethamine. Compounds 20 through 30 showed a similarity score to pyrimethamine of 0.62 to 0.65. The greater the similarity score, the higher the correlation to pyrimethamine. Compounds 20 through 31 have a very broad variation in molecular scaffolding, however, the molecular properties are still showing zero violations of RO5, indicating favorable drug-likeness to bioavailability.

Compounds 20 to 31 show aromatic rings, hetero-atomic aromatic rings, halogen atoms, and aliphatic substituents. Overall the variation in structure is broad, however consistency in molecular properties with the features of pyrimethamine is clear and paramount.

Looking at the whole of Table 1, the molecular properties, the Grubb's test for outliers showed no outliers for all 31 agents in the following descriptors: Log P, polar surface area, number of atoms, molecular weight, number of oxygen & nitrogen atoms, number of –OH and –NH_n groups, number of rotatable bonds, and molecular volume. This result is a clear indication that the six isosteres and remaining 24 compounds are extremely consistent in properties. This and the favorable indicator for favorable bioavailability supports the contention that these 30 novel drug designs would have efficacy of biological activity analogous to pyrimethamine and potential clinical application.

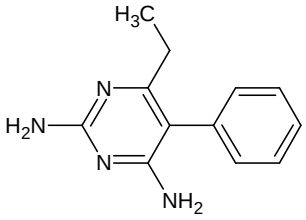
3.2 Pattern Recognition Identification of Interrelationships

Structural similarity is one of the basic underlying principles in drug discovery and development [16]. The basic underlying principle of compound subset selection and structure-activity relationship studies presumes that compounds which are similar to biologically active molecules should themselves also be active, and vice versa [16]. Therefore, determination of similarity is among the most intensively studied areas with numerous methods published [16].

In data mining, hierarchical cluster analysis is a method of cluster analysis that seeks to build a hierarchy of clusters [17,18]. Cluster Analysis goals are to relate the grouping a collection of objects (i.e., observations, individuals, cases, or data rows) into subsets or clusters, where those within each cluster are more closely related to one another than objects assigned to different clusters [17,18]. Central to cluster analysis is the concept of degree of similarity (or dissimilarity) between the individual objects being clustered. There are two major methods of clustering: hierarchical clustering and K-means clustering. Both will be applied to this study.

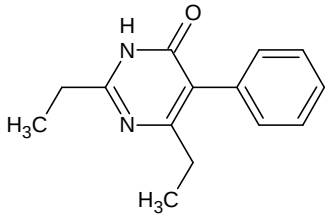
The Dendrogram presented in Fig. 4 shows the results of hierarchical cluster analysis of all 31 agents presented in Table 1. Conditions of cluster analysis are Euclidean distance and single-linkage clustering (distance between one cluster and another cluster is equal to the shortest distance from any member of one cluster to any member of the other cluster) [17, 18]. The Dendrogram is divisive approach with all observations starting in one cluster (A), and then splitting as it moves down the hierarchy.

8



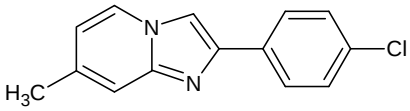
CCc1nc(N)nc(N)c1c2ccccc2

9



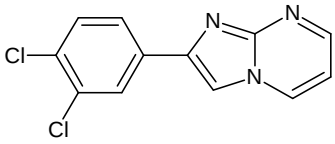
O=C2NC(=NC(CC)=C2c1ccccc1)CC

10



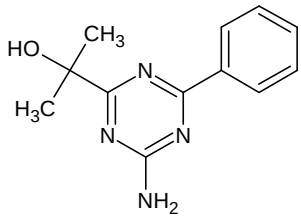
Clc1ccc(cc1)c2cn3ccc(C)cc3n2

11



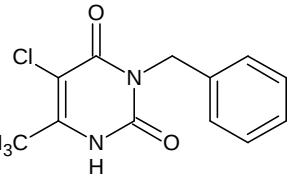
Clc1ccc(cc1Cl)c2cn3cccnc3n2

12



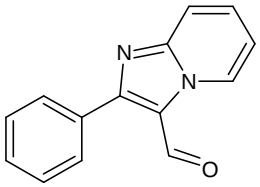
Nc1nc(nc(n1)C(C)(C)O)c2ccccc2

13



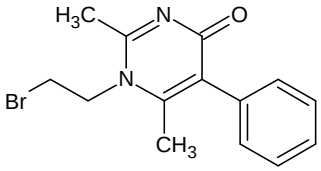
O=C2C(Cl)=C(C)NC(=O)N2Cc1ccccc1

14



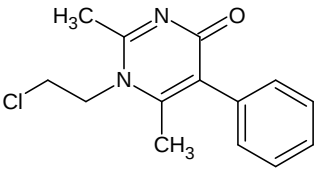
O=Cc1c(nc2ccccn12)c3ccccc3

15



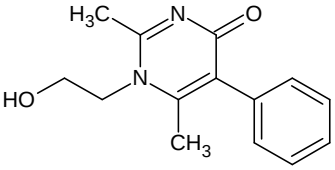
BrCCN1C(C)=C(C(=O)N=C1C)c2ccccc2

16



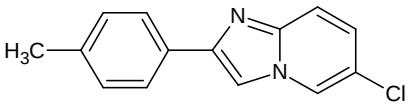
ClCCN1C(C)=C(C(=O)N=C1C)c2ccccc2

17



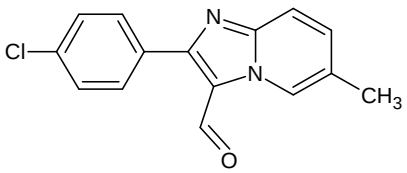
OCCN1C(C)=C(C(=O)N=C1C)c2ccccc2

18



Cc1ccc(cc1)c2cn3cc(Cl)ccc3n2

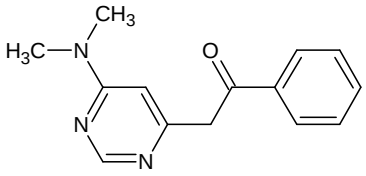
19



Clc1ccc(cc1)c2nc3ccc(C)cn3c2C=O

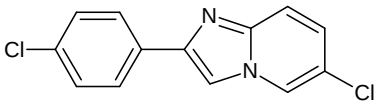
Fig. 2. Outcome of heuristic similarity search with compounds having similarity score 0.66 to 0.87

20



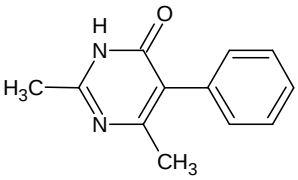
O=C(Cc1cc(ncn1)N(C)C)c2ccccc2

21



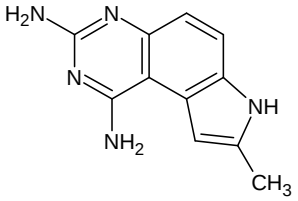
Clc1ccc(cc1)c2cn3cc(Cl)ccc3n2

22



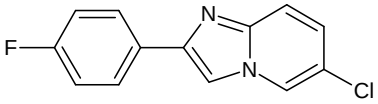
O=C2NC(C)=NC(C)=C2c1ccccc1

23



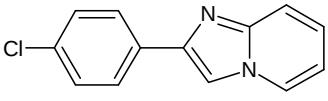
Cc3cc2c(ccc1nc(N)nc(N)c12)n3

24



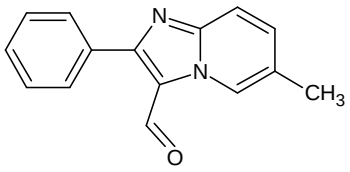
Fc1ccc(cc1)c2cn3cc(Cl)ccc3n2

25



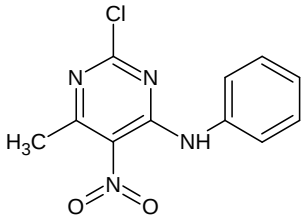
Clc1ccc(cc1)c2cn3ccccc3n2

26



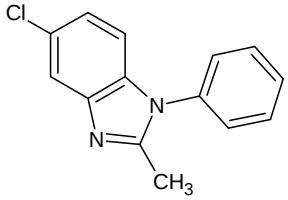
Cc1ccc2nc(c(C=O)n2c1)c3ccccc3

27



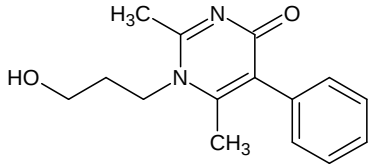
Cc2nc(Cl)nc(Nc1ccccc1)c2N(=O)=O

28



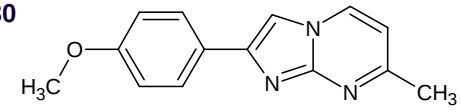
Clc1cc2nc(C)n(c2cc1)c3ccccc3

29



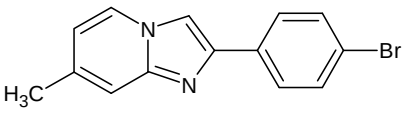
CC1=C(C(=O)N=C(C)N1CCCO)c2ccccc2

30



COc1ccc(cc1)c2cn3ccc(C)nc3n2

31



Brc1ccc(cc1)c2cn3ccc(C)cc3n2

Fig. 3. Outcome of heuristic similarity search for compounds having similarity scores 0.62 to 0.65

In divisive cluster analysis, all objects (compounds) begin in cluster A, but then are resolved into clusters holding analogous agents steadily until all 31 agents are distinguished. Leading from node A, agent 6 (isostere) is determined to be unique from the remaining 30 agents, most likely due to the highest polar surface area (103.85 Angstroms^2). Agent 6, 27,

22, and 23 are determined to be distinguished from all other agents found under node B. Agent 19 is determined to be distinct from all other agents found under node B, potentially due to a very high Log P (4.08), but with a low polar surface area (34.38 Angstroms^2). Virtually all the remaining agents are found under node C, leading to node D containing agents 8, 12, 5, 7, 4, 3, 2, and 1 (pyrimethamine). Ultimately, the agents most similar to pyrimethamine (agent 1) are the isosteres 2, 3, 4, and 7. This outcome is accordant that isostere compounds are molecules or ions having the same number of atoms and/or the same number of valence electrons.

K-means cluster analysis an unsupervised algorithm to solve the clustering goal. The method classifies a data set to a certain number of clusters determined prior to analysis [18]. Again, the outcome are a number of clusters chosen before analysis where each drug is designated with other drugs that have the highest similarity [18]. The result of K-means cluster analysis of agents shown in Table 1 are following, with isosteres represented in bold:

Cluster 1: 11, 19, 21, and 31.
Cluster 2: 13, 16, 17, 20, and 29.
Cluster 3: 1 (pyrimethamine), 2, 3, 4, 5, 6, 7, 8, 12, 23, and 27.
Cluster 4: 9, 10, 14, 18, 22, 24, 25, 26, 28, and 30.
Cluster 5: 15.

Interestingly, all isosteres found in search engine are determined to have the highest similarity to the parent compound pyrimethamine (cluster 3). This result is consistent with hierarchical cluster analysis. In addition, agents 8, 12, 23, and 27 are added to cluster 3. Compound 15 (cluster 5) is determined to be distinct from all other compounds. The determination of isosteres of pyrimethamine shows to be of the highest similarity among 30 compounds selected by search engine is consistent for both approaches to cluster analysis (hierarchical and K-means).

Box plots or box and whisker diagrams is a standardized statistical method to visualize the displaying the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum [18]. Box and whisker plots are very useful when large numbers of observations are involved and when two or more data sets are being compared. Boxplots display two common measures of the

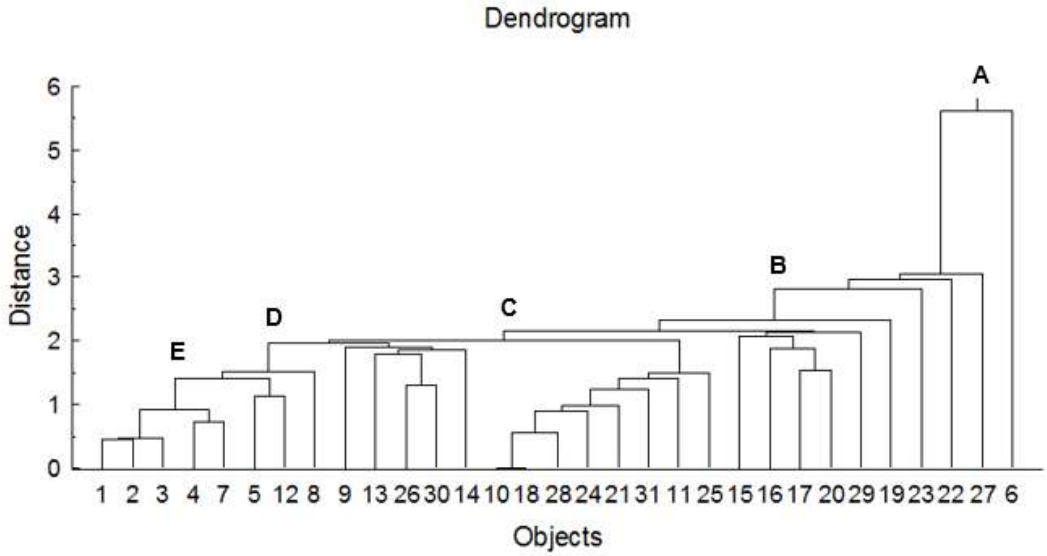


Fig. 4. Hierarchical cluster analysis of compound molecular properties presented in Table 1. Analysis utilized standard Euclidean distance with single linkage for clusters. Super-node A, encompassing all compounds, is divided into clusters of compound (see Table 1) based on highest similarity. Node B excludes compounds 22, 27, and 6. Node C excludes 15, 16, 17, 20, 29, 19, 23, 22, 27, and 6. Node D includes compounds 1, 2, 3, 4, 7, 5, 12, and 8. Node E includes compounds 1, 2, 3, 4, and 7

variability or spread in a data set: Range the spread of all the data; and Interquartile range (the middle half of a data set falls within the interquartile range).

Observing the box plots for molecular properties (see Fig. 5), the medians are indicated within the box with the range by the whiskers [18]. The box plots show visually the relative position of the medians and ranges of each molecular property listed in Table 1. The broadest ranges (maximum to minimum) are found with polar surface area, molecular weight, and volume. Their ranges, respectively, are 17.31 Angsroms^2 to 103.85 Angstroms^2, 200.24 Dalton to 307.19 Dalton, and 187.84 Angstroms^3 to 247.19 Angstroms^3.

Interestingly, most molecular properties have a very small range from minimum to maximum that are very close to the median. These descriptors are Log P, number of atoms, number of oxygen & nitrogen atoms, number of –OH & -NH_n groups, violations of the rule of 5, and number of rotatable bonds. The oxygen, nitrogen, hydroxyl (-OH), and amine (-NH_n) groups account for the hydrogen bond acceptor and hydrogen bond donor characteristic of the molecule. These values, shown in the box plots, are highly consistent for all 31 agents and very small in range (i.e. having whiskers very close to the medians).

Utilizing box plots presented in Fig. 5, it is readily seen that numerical values from minimum to maximum is small and are extremely close to the median, upper, and lower quartile (except for PSA, molecular weight, and molecular volume). Most of the descriptors of all 30 compounds are highly consistent, numerically close in proximity, and very close to the group medians. In addition, and further evidence the 31 compounds have very strong positive relationship are results of Pearson correlation of one compound to another (applying all properties, see Table 1), are all greater than 0.9700. Taken together, the box plots and Pearson r correlation clearly show that pyrimethamine, isosteres, and similarity search structures are very highly similar and with properties indicating clinically favorable drug-likeness, permeability, and membrane penetration.

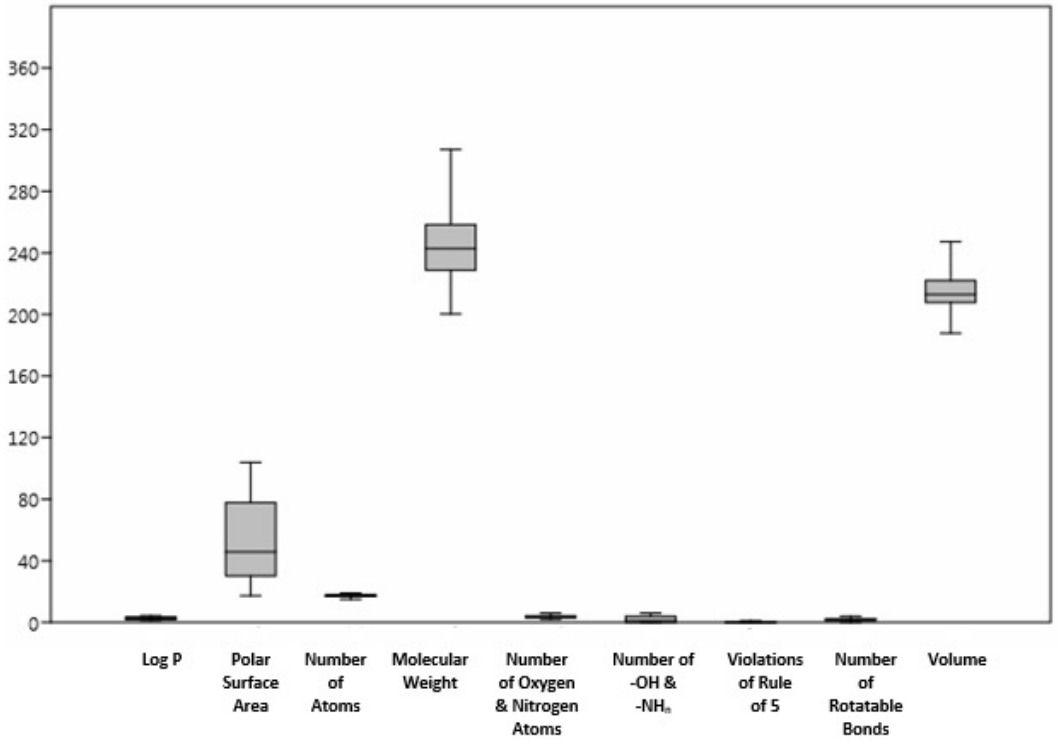


Fig. 5. Box plots of properties presented in Table 1, with whiskers of one standard deviation
This representation shows numerical dispersion of values and range of numerical values

3.3 Multiple Regression and Prediction

Multiple regression analysis is a powerful technique for predicting the value of an dependent variable from the known value of two or more independent variables, referred to as predictors [19]. In this analysis, the predictors are molecular properties such as Log P and polar surface area, having significance as pharmaceutical properties. For compounds 1 to 31, and utilizing the properties given in Table 1, we can form analysis with Log P, polar surface area (PSA), number of atoms (nAtoms), number of oxygen & nitrogen atoms (nON), number of –OH and $–NH_n$, molecular volume (MV), and number of rotatable bonds (nRotB):

MW = -85.632 + 11.026(LogP) – 0.1324(PSA) – 9.156(nAtoms) + 14.362(nON) – 5.014(nOHNH) + 5.445 (rule of 5) – 15.523(nRotB) + 2.107(MV) (1)

The relationship (1) shows an R^2 of 0.7060, indicating the model accounts for 70.60% of the variance in MW based on the model. The dependent variables of greatest significance are the Log P (*P*=.02), number of rotatable bonds (*P=.04*), and molecular volume (*P*=.01).

Various techniques disseminated to find better medicines faster include QSAR (Quantitative structure–activity relationship), structure based techniques, homology modelling (generating a homologous structure to the parent drug), and high throughput screening [20]. This study has demonstrated the efficacy of applying isosteres and similarity search utilizing heuristic search engine. Isosteres for lead modification has been shown to be highly successful approach to modify the activity of a lead drug (pyrimethamine here) and modify toxicity [21]. Applying isosteres in drug design has been highly successful, in part due to the only subtle changes of size, shape, Log P, and hydrogen bonding [21]. The use of in-silico chemistry and molecular modeling has great momentum and potential [22]. In-silico methods will continue to have substantial importance in prediction of novel drugs [22].

This study has shown that specific isostere substitution and rational similarity search produced 30 potential agents for anti-protozoan usage. The drug pyrimethemine has been shown to be effective for treatment of protozoa diseases. The 30 agents presented here have favorable drug-likeness based on the rule of 5 (i.e. good bioavailability and membrane permeation [14]). Their properties are modified only slightly and be statistical analysis are shown to adhere to rigid analogy to parent drug pyrimethemine. With the great success of isostere modification of a successful lead, these 30 agents (2 to 31) have potential to be useful in the clinical treatment of protozoan diseases.

4. CONCLUSION

Pyrimethamine is an important drug for the treatment of protozoan infections. Thirty agents were identified to be analogous to pyrimethamine following heuristic search engine, six of the agents as isosteres. The result of search engine identified 30 agents that were highly consistent in important pharmacological properties such as Log P and polar surface area. Grubb's test for outliers showed no outlying values for Log P, polar surface area, molecular weight, and four other molecular properties. All compounds in this study showed favorable bioavailability, having zero violations of the rule of 5. Hierarchical cluster analysis and K-means cluster analysis showed that the isostere compounds were most similar to pyrimethamine. Multiple regression analysis produced a mathematical model for predicting properties of compounds of similar medicinal use. The study and design of new medicinal agents for treatment of protozoan infections is a vital endeavor with great demand in the future.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

ACKNOWLEDGEMENTS

This study was supported by the College of Arts & Sciences, University of Nebraska at Omaha, Omaha Nebraska USA.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Darde ML, Ajzenberg D, Smith J. *Toxooplasma gondii*: The model apicomplexan. Perspectives and methods. London: Academic Press/Elsevier; 2011.

2. Flegr J, Prandota J, Sovickova M, Israili ZH. Toxoplasmosis - A global threat. Correlation of latent toxoplasmosis with specific disease burden in a set of 88 countries. PLoS ONE. 2014;9(3):e90203.
3. Pappas G, Roussos N, Falagas ME. Toxoplasmosis snapshots: Global status of *Toxoplasma gondii* seroprevalence and implications for pregnancy and congenital toxoplasmosis. Int J Parasitol. 2009;39(12):1385-94.
4. Berdoy M, Webster J, Macdonald D. Fatal attraction in rats infected with *Toxoplasma gondii*. Proc Biol Science. 2000;267(1452): 1591-94.
5. Webster JP, Kaushik M, Bristow GC, McConkey GA. *Toxoplasma gondii* infection, from predation to schizophrenia: Can animal behavior help us understand human behavior? J Exp Biol. 2013; 216(Pt 1):99-112.
6. Hari Dass SA, Vyas A. *Toxoplasma gondii* infection reduces predator aversion in rates through epigenetic modulation in the host medial amygdala. Mol Ecol. 2014;23(24):6114-22.
7. Alcamo E, Warner JM. Schaum's outline of microbiology. New York: McGraw Hill; 2009.
8. Keen EC. Beyond phage therapy: Virotherapy of protozoal diseases. Future Microbiology. 2013;8(7):821-23.
9. Trevor AJ, Katzung BG, Masters SB. Katzung & Trevor's Pharmacology: Examination & Board Review. New York: McGraw-Hill Professional; 2007.
10. Anonymous. Anti-protozoal drugs. Pharmaceutical Drug Manufacturers. Available:http://www.pharmaceutical-drug-anufacturers.com/pharmaceutical-drugs/anti-protozoal-drugs.html (Accessed 17th February 2017)
11. Rossi S. Australian medicines handbook. Rundle Mall: Australian Medicines Handbook Pty; 2013.
12. Torgerson PH, Macpherson DNL. The socioeconomic burden of parasitic zoonosis: Global trends. Vet Parasitol. 2011;182:79-95.
13. Clark DE. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. J Pharm Sci. 1999;88(8):807-14.
14. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997;23:3-25.
15. Palm K, Stenberg P, Luthman K, Artursson P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharm Res. 1997;14(5):568-71.
16. Schwaha R, Ecker GF. The similarity principle-new trends and applications in ligand-based drug discovery and ADMET profiling. Sci Pharm. 2008;76:5-18.
17. D'andrade R. U-Statistic Hierarchical Clustering. Psychometrika. 1978;4:58-67.
18. Legendre P, Legendre L. Numerical ecology. 2nd Ed. Amsterdam: Elsevier; 1998.
19. Pedhazur EJ. Multiple regression in behavioral research. Orlando: Harcourt Brace; 1997.
20. Seddon G, Lounnas V, McGuire R, van den Bergh T, Bywater RP, Oliveira L, Vriend G. Drug design for ever, from hype to hope. J Comput Aided Mol Des. 2012;26(1):137-50.
21. Silverman RB. The organic chemistry of drug design and drug action. San Diego: Academic Press; 1992.
22. Wadood A, Ahmed N, Shah L, Ahmad A, Hassan H, Shams S. *In-silico* drug design: An approach which revolutionarised the drug discovery process. OA Drug Design & Delivery. 2013;1(1):3-7.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/19021