

How Far Away From Nature are We? Analysis of Correlation Similarities between Descriptors of the Drug Bank and Tripeptides Molecules

Eitner K¹, Gawęda T¹, Koch U²

¹Adam Mickiewicz University, Faculty of Chemistry, 89b Umultowska Str., PL-61-614 Poznań, Poland

²Lead Discovery Center, Emil-Figge-Str. 76a, 44227 Dortmund, Germany

Abstract

A Statistical analysis was performed of similarities between 2D topological descriptors from the Drug Bank molecule database and 8,000 tripeptides (all possible amino acid combinations encoded by nucleic acids). The correlation between theoretically calculated properties of tripeptide molecules (MW, AlogP, Topological PSA, hydrogen bond donors and hydrogen bond acceptors) and topological descriptors from Drug Bank showed major similarities between simple tripeptides and compounds with dedicated bioactivity developed in laboratories. The paper presents histograms for the distribution of the number of compounds with similar molecule properties as encoded by the descriptors. A simple, innovative methodology for the large scale analysis of statistical data and their correlation has been developed within our study. Some hypothesis indicates that highly processed food is a natural antibacterial and antiviral barrier. Our research in comparison with literature data proves that many xenobiotics are topologically similar to natural metabolites being tripeptides and some have similar therapeutic applications.

Keywords: Tripeptides; Drugs; Inhibition; QSAR; Statistical analysis; Descriptors

Introduction

The twenty standard amino acids and their peptide protein-forming combinations represent in a unique way a complete range of molecular interactions, molar masses, solubility in aqueous and non-polar solutions, surface properties and a number of other features of chemical compounds. Amino acids and polypeptides, their low-molecular weight compounds, are distributed through every living organism as components of various molecules from structural ones to reaction catalysts and metabolites. They all have evolutionary usefulness and bioactivity resulting from environmental factors. Food is for animals an important source of amino acids and peptides.

According to some theories [1-3] food processing by hominids contributed to the rapid evolutionary brain development. Is it therefore possible that metabolic peptide concentrations protect us naturally against pathogenic microorganisms, their metabolites and proteins? Are the folk beliefs in the beneficial power of bouillon and other foods with highly hydrolysed protein justified? Most importantly, how to demonstrate in the simplest way possible therapeutic, physical and chemical similarities?

Humans have always pursued their need for helping and saving health and life at risk and used mixtures and substances in whose therapeutic power they believed.

The present-day pharmaceutical industry has developed thousands of compounds with defined bioactivity in order to inhibit many bacterial and viral proteins. The compounds compete for active sites with metabolites, thus inducing the desired therapeutic effect. It then seems natural to ask how similar and how different artificial inhibitors are from natural polypeptides. QSAR analysis is one of the current methods for the comparison of and search for compounds with desired therapeutic activity. Using QSAR, we can characterize and compare many compounds and select potential inhibitors in a cost-effective manner. 8000 tripeptides were selected within the present study, that is, all the possible combinations of amino acids encoded by nucleic acids, and a drug database with 4886 chemical compounds: >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/

peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs; (data for May 2010). We expect that through the comparison of tripeptides and drugs, it will be possible to determine similarities which affect their therapeutic efficacy. We found that several xenobiotics and tripeptides characterized in our calculations have a well-defined and similar therapeutic applications [4-7].

Experimental Procedures

The manuscript uses descriptor analysis (MW, AlogP, Topological PSA, hydrogen bond donors and hydrogen bond acceptors) generated by PADEL-Descriptor software (<http://padel.nus.edu.sg/software/padeldescriptor/>) for the Drug Bank database [8, 9] and for 8000 tripeptides (combinations of 20 amino acids). The tripeptide database was generated in the SMILE format using the Chem Axon Software molconvert application. Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin 2.5.1, 2009, ChemAxon, <http://www.chemaxon.com>.)

Another stage was to transfer data to the MySQL database using our proprietary Perl programme to facilitate efficient selection and data analysis on a further stage. Furthermore, statistical analysis between selected DrugBank descriptors and tripeptides was possible.

The data from the database were grouped by value ranges with defined accuracy; subsequently, descriptor value histograms were generated separately for the Drug Bank compounds and tripeptides. (Figures 1, 2).

***Corresponding author:** Krystian E, Adam Mickiewicz University, Faculty of Chemistry, 89b Umultowska Str, PL-61-614 Poznań, Poland, Tel: 0048 513078981; E-mail: eitner@amu.edu.pl

Received July 21, 2014; **Accepted** September 29, 2014; **Published** October 08, 2014

Citation: Eitner K, Gawęda T, Koch U (2014) How Far Away From Nature are We? Analysis of Correlation Similarities between Descriptors of the Drug Bank and Tripeptides Molecules. J Theor Comput Sci 1: 220. doi: [10.4172/2376-130X.1000120](https://doi.org/10.4172/2376-130X.1000120)

Copyright: © 2014 Eitner K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

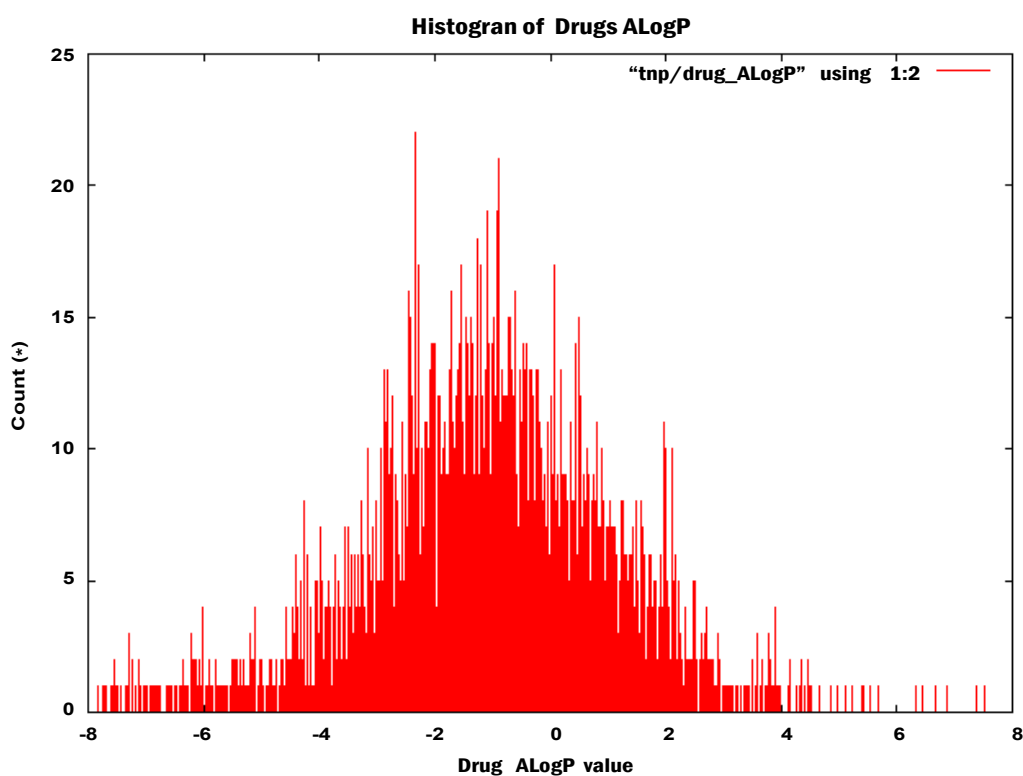


Figure 1: Histogram of AlogP distribution for all 8000 of considered tripeptides.

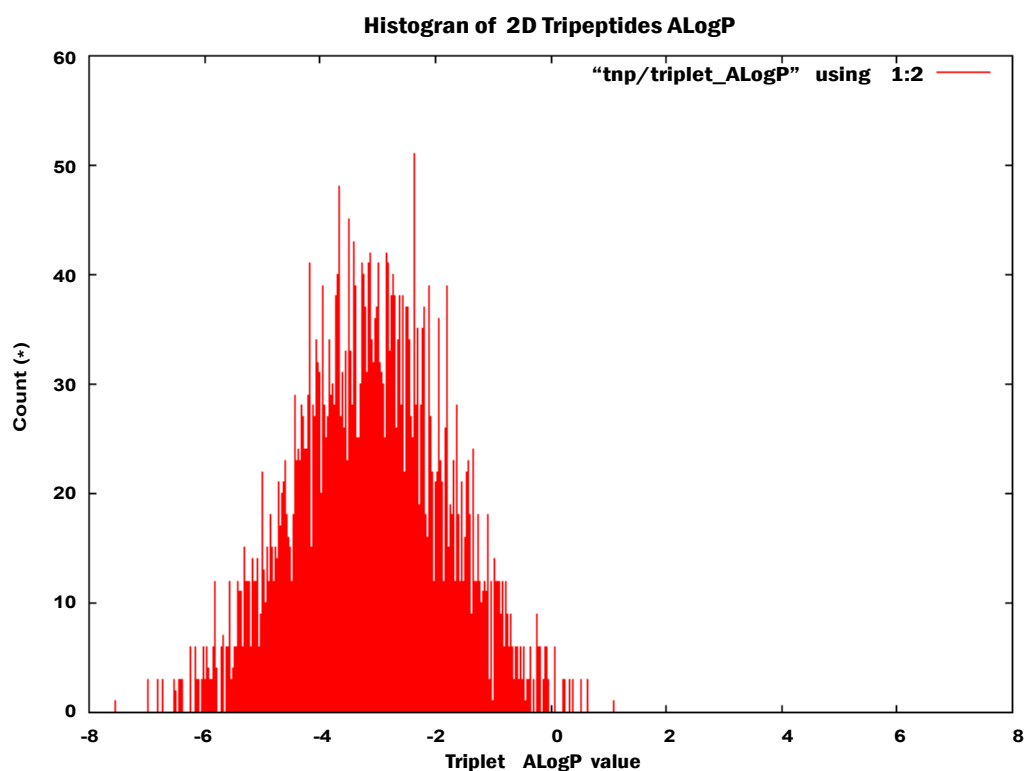


Figure 2: Histogram of AlogP distribution for all 4886 Drug Bank molecules.

Values from both databases were correlated by calculating a difference matrix between Drug Bank and tripeptides. The similarity matrix contains 39088000 similarity values for each of the five individual descriptors (4886 molecules from “All Drug Structures” of Drug Bank *8,000 tripeptides).

Based on the lowest descriptor value differences between the compounds in both databases, we selected the most similar compounds for further QSAR 2D analysis (Figure 3).

The range of differences determining similarity values within limits according to the Lipiński rule or within a wide value range for tripeptides was determined from histograms generated previously for each descriptor.

The value range for the selected descriptors used for the selection of compounds with the highest similarity (highest topological similarity) is shown in (Table 1). The last stage of our calculation involved the generation of intersection for a group with the highest similarity among Drug Bank compounds and tripeptides specified by selected descriptors.

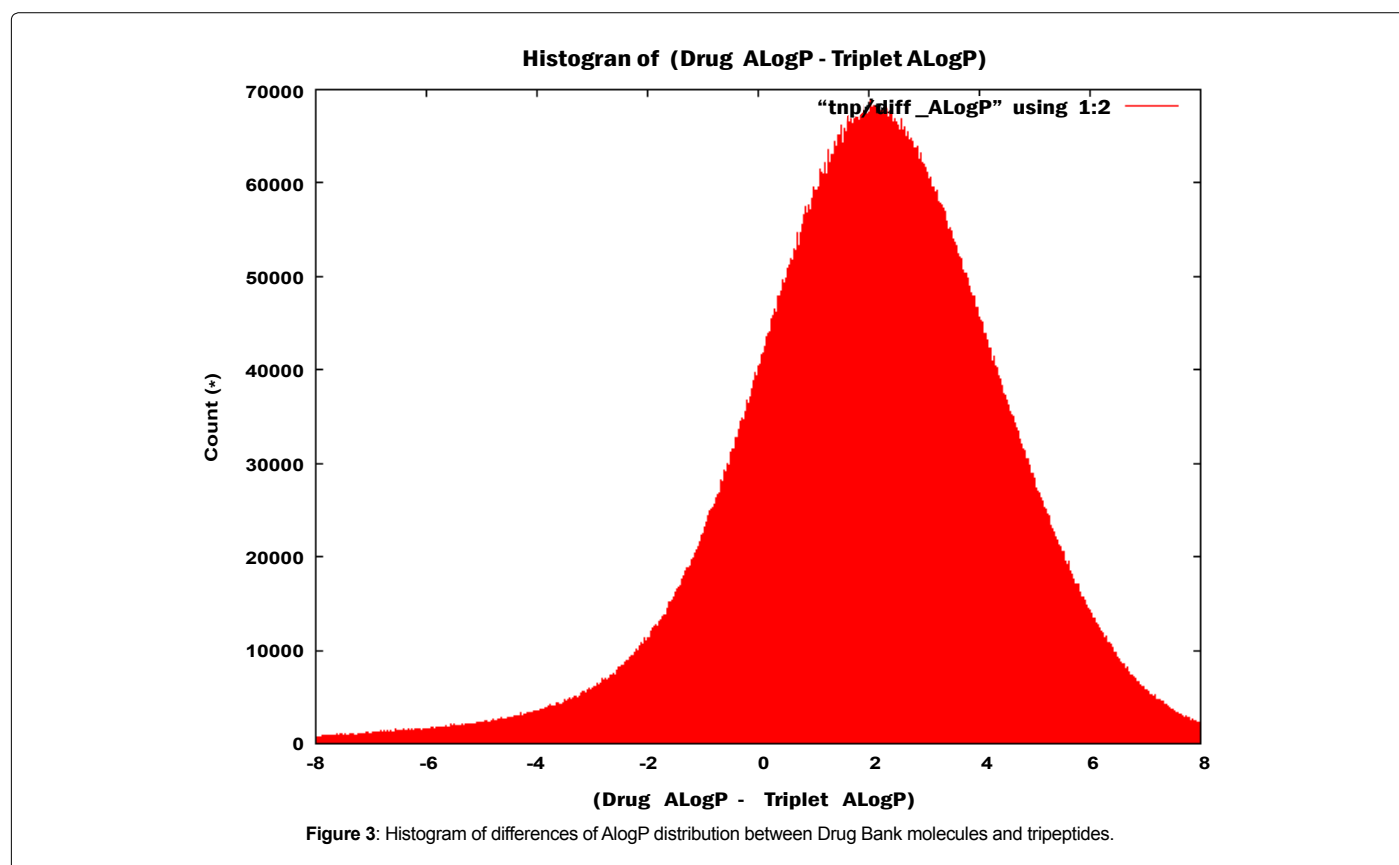
Results Analysis and Discussion

The AlogP values for 95% of tripeptides are within a range of -5.43, -0.76 (median value -3.10). According to the Lipiński rule, distribution ratio values for drugs are below 5. As for the Drug Bank compounds, 95% compounds are within a range of AlogP values of between -6.90 and 2.60 (median value -2.15). The values prove higher contribution of polar compounds in the xenobiotic group than among tripeptides. This results from the fact that drugs should have high solubility, and

we know that many of them are taken orally and absorbed from the gastrointestinal tract.

The analysis of TPSA calculations shows that drugs show a much wider distribution of values than the tripeptides. 27% of drugs are within a range for all the tripeptides; however, all tripeptides are within a range for the drugs. By assuming a criterion of difference in TPSA values according to (Table 1), 1.14% correlation between drugs and tripeptides was obtained. It follows from the calculation that statistically Drug Bank drugs are more polar than tripeptides. This is consistent with the AlogP calculations.

As for pharmacokinetics, TPSA [10] and AlogP are highly important for the determination of ADME (Absorption, Distribution, Metabolism, and Excretion) parameters which describe xenobiotic behavior in the body. It follows from the calculation that the distribution ratio values for all the 8000 tripeptides are within a range determined for drugs. 68.83% (3,363/4,886) of Drug Bank compounds have molar masses within 160-500 [11], while 99.11% tripeptides are within this range (7,929/8,000). It follows from the analysis of molar mass differences between the compounds in the databases that 3.94% correlations, or 1,539,329 of 39,088,000 possible similarities between drugs and tripeptides, are within a difference range of <-10; 10 a.m.u.>. However, the molar mass criterion refers only to low-molecular weight compounds, being pointless when searching for correlations between small tripeptides and compounds with extended structures (polysaccharides or large polypeptides). For simple chemical substances, molar mass depends on the number of non-hydrogen atoms in the molecule and only combined with selected atom classes (number of hydrogen bond acceptor or donor atoms), it provides a consistent and qualitative picture of molecules.



84% drugs fulfill the Lipinski rule (not more than 10 hydrogen bond acceptors) and 96% have not more than 5 hydrogen bond donors.

As for tripeptides, the values are 69% for hydrogen bond acceptors and 42% for donors.

The phrase “not more than 10 hydrogen bond acceptors” refers to the process of inhibitor molecule seizing by protein molecules. “Not more than 5 hydrogen bond donors” contributes to the sticking of molecules to active sites and/or protein surface. Hydrogen bonds are the most vital group of protein-ligand interactions; however, large accumulation of acceptors and donors on one molecule is not favorable for the transport of the system which strongly interacts with its environment. It follows from our calculation that the Drug Bank compounds are more consistent with the Lipinski rule. It is noted that being natural metabolites, tripeptides are metabolized more rapidly and transported more easily; from an evolutionary perspective they are thus likely to have higher potential for interacting with other metabolites and proteins.

Correlation between tripeptides and drugs shows that the correlation value is 12.37% at a level of one acceptor and one donor differences (4,833,500 correlations between the databases) for acceptors and 4.49% (1,754,086 correlations between the databases) for donors.

Through generation of intersection which would fulfil all the criteria defined by the value ranges from Table 1, we found that 163 Drug Bank compounds are topologically similar with 1,617 tripeptides (Table S1, Supplemental Material). Even though 263,571 similarities between drugs and tripeptides is not much (0.67% of all possible similarities), it is obvious that competing metabolites (tripeptides) exist for many drugs absorbed into the body.

The most important conclusion from our calculation is to note that xenobiotics introduced into the body compete with natural metabolites (tripeptides in this case). It is expected that this may be one of the natural defense mechanisms against toxins in the body. Competition for active sites of bacterial and viral proteins is a natural consequence of the similarities between the small molecules from Drug Bank and natural tripeptides.

A comparison of GGL tripeptide (commercial name: Diapine) and DB00428 compound (streptozotocin) shows that both compounds stimulate blood glucose level by increasing insulin concentration in male diabetes II [4,5]. Likewise, tripeptide GHK and DB05475 (gamma-D-glutamyl-L-tryptophan) have demonstrated efficacy in treating various viral and bacterial infections [6,7] (See supplementary material).

Descriptor	Value range	Number of similarities/%
MW	<-10; 10>	1539329 / 3.94
AlogP	<0.52; 3.52>	18608726 / 47.61
TopoPSA	<-5.0; 5.0>	446254 / 1.14
nHBAcc	<-1; 1>	4833500 / 12.37
nHBDon	<-1; 1>	1754086 / 4.49

MW – Molecular Weight
AlogP – Atomic Based Partition Coefficient
TopoPSA – Topological Polar Surface Area
nHBAcc – Number Of Hydrogen Bond Acceptors
nHBDon – Number Of Hydrogen Bond Donors
100% similarities means 39088000 similarity values for each of the five individual descriptors.

Table 1: Difference value ranges for descriptor similarity in the correlation between the tests compounds, used for determining the intersection of both compound databases (Drug-Bank vs. tripeptides)

It is obvious that many drugs might be replaced by proper diet, or in other words, proper diet supports the treatment of many bacterial and viral diseases [12].

The calculation results and literature analysis confirms the hypothesis that highly processed food is a natural antibacterial and antiviral barrier, the first line of body defense which has ensured evolutionary development of human beings [2].

Summary

The paper presents results of calculation of topological descriptor value correlation between the Drug Bank database (chemical compounds with known bioactivity) and all the possible tripeptide molecules composed of the 20 amino acids encoded by nucleic acids. Proprietary software for the analysis of extensive computational data was developed to generate histograms of descriptor distribution for the molecules from both databases. This paper shows that similarities between selected descriptors include the whole range of values seen for tripeptides and drugs. In other words, drugs used in pharmacy have all the features of tripeptides: similar molar masses, significant similarity in the number of donor and acceptor atoms, AlogP range (albeit broader for the Drug Bank compounds) and Topological PSA.

The first direct conclusion from our calculation is that highly processed food may be a source of high concentrations of tripeptides introduced into the body which may compete with artificially developed drugs for protein active sites.

As the intersection of the properties under the descriptors between both databases was determined, a large group of drugs whose molecules are highly similar to tripeptides in the defined descriptor value range was revealed. The resulting intersection which fulfils the criteria (Table S1) contains 263 571 similarities between the Drug Bank compounds and tripeptides. It is noted that a number of Drug Bank compounds are similar to more than one tripeptide and many tripeptides are similar to more than one drug. In particular, 1617 tripeptides are similar to 163 drugs. This results in much higher competitiveness of tripeptides than drugs at much higher metabolic concentrations. On the other hand, drug therapeutic concentrations are much lower than tripeptide concentrations. The present study shows that the folk beliefs that bouillon is one of the most efficient “drugs” which support cold or flu treatment are not without a reason.

Acknowledgment

KE would like to thank the Foundation for Polish Science for the support through the FOCUS programme.

Supporting Information Available

Histograms of descriptor values distribution and list of Drug Bank compounds correlated with tripeptides are available in Supporting Information.

References

1. Kaplan H, K Hill, J Lancaster, Hurtado AM (2000) a theory of human life history evolution: diet, intelligence, and longevity. *Evolutionary Anthropology* 9:156-185.
2. Milton K (1999) A hypothesis to explain the role of meat-eating in human evolution. *Evolutionary Anthropology* 8: 11-21.
3. Gibbons A (2007) Paleoanthropology: Food for Thought. *Science* 316: 1558-1560.
4. Zhang J, Xue C, Zhu T, Vivekanandan A, Pennathur S, et al. (2013) A Tripeptide Diapine Effectively Lowers Blood Glucose Levels in Male Type 2 Diabetes Mice by Increasing Blood Levels of Insulin and GLP-1. *PLoS ONE*: e83509.

5. Wang Z, Gleichmann H (1998) GLUT2 in pancreatic islets: crucial target molecule in diabetes induced with multiple low doses of streptozotocin in mice. *Diabetes* 47: 50-56.
6. Meiners S, Eickelberg O (2012) Next-generation personalized drug discovery: the tripeptide GHK hits center stage in chronic obstructive pulmonary disease. *Genome Medicine* 4: 70.
7. Suzuki H, Kato K, Kumagai H (2004) Development of an efficient enzymatic production of gamma-D-glutamyl-L-tryptophan (SCV-07), a prospective medicine for tuberculosis, with bacterial gamma-glutamyltranspeptidase. *J Biotechnol.*, 5: 291-295.
8. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. (2008) a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 36: D901-906.
9. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, et al. (2006) a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 1: D668-672.
10. Ertl P, Rohde B, Selzer P (2000) Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J Med Chem* 43: 3714-3717.
11. Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. *J CombinChem* 1: 55-68.
12. Eng C, Kramer KC, Zinman B, Retnakaran R (2014) Glucagon-like peptide-1 receptor agonist and basal insulin combination treatment for the management of type 2 diabetes: a systematic review and meta-analysis, *The Lancet*.