

TOXICITY VS POTENCY: ELUCIDATION OF TOXICITY PROPERTIES DISCRIMINATING BETWEEN TOXINS, DRUGS, AND NATURAL COMPOUNDS

SWANTJE STRUCK^{1†}
swantje.struck@charite.de

ULRIKE SCHMIDT^{1†}
ulrike.schmidt@charite.de

BJOERN GRUENING¹
bjoern@gruenings.eu

INES S. JAEGER^{1,2}
ines.jaeger@charite.de

JULIA HOSSBACH¹
julia.hossbach@charite.de

ROBERT PREISSNER^{1*}
robert.preissner@charite.de

¹ *Structural Bioinformatics Group, Institute of Molecular Biology and Bioinformatics
Charité-University Medicine, Arnimallee 22, 14195 Berlin, Germany*

² *Department of Cardiology and Angiology, Charité-University Medicine,
Schumannstr. 20/21, 10117 Berlin, Germany*

Within our everyday life we are confronted with a variety of toxic substances. A number of these compounds are already used as lead structures for the development of new drugs, but the amount of toxic substances is still a rich resource of new bioactive compounds. During the identification and development of new potential drugs, risk estimation of health hazards is an essential and topical subject in pharmaceutical industry. To face this challenge, an extensive investigation of known toxic compounds is going to be helpful to estimate the toxicity of potential drugs. "Toxicity properties" found during those investigations will also function as a guideline for the toxicological classification of other unknown substances.

We have compiled a dataset of approximately 50,000 toxic compounds from literature and web sources. All compounds were classified according to their toxicity. During this study the collection of toxic compounds was investigated extensively regarding their chemical, functional, and structural properties and compared with a dataset of drugs and natural compounds.

We were able to identify differences in properties within the toxic compounds as well as in comparison to drugs and natural compounds. These properties include molecular weight, hydrogen bond donors and acceptors, and functional groups which can be regarded as "toxicity properties", i.e. attributes defining toxicity.

Keywords: toxins, drugs, natural compounds, toxicity

* corresponding author

† equally contributing authors

1. Introduction

Toxins are hazardous substances causing illness or damage to an exposed organism when inhaled, swallowed or absorbed through the skin. The famous physician Paracelsus

(1493-1541) already mentioned: “Dosis sola venenum facit” (“Only the dose makes the poison”) which is still a central concept of toxicology. This dose-dependency implicates that even water may evoke toxic effects when given in high amounts, and, on the contrary, small doses of a powerful toxin may lead to healing.

Toxins constitute a very diverse group of substances, ranging from enzymes up to small chemical compounds affecting just as many different targets [1]. It is of great scientific interest to classify toxins and to compare their toxic effects in order to identify new toxins and to understand the biological mechanisms they are involved in. There are different measurements to estimate toxicity: LD50 and LC50 (lethal dose or concentration at which 50% of a population dies) are widely established, but also TGI (total growth inhibition), NOEL (no observable effects limit), or LOEL (lowest observable effects level) are used.

In recent publications it was considered to perform QSAR (Quantitative Structure-Activity Relationship) or QSPR (Quantitative Structure-Property Relationship) analyses to predict the toxicity based on chemical and physical properties without further experimental investigations [2,3]. The motivation for the study at hand was the collection of a dataset of structures with corresponding toxicity information which formed the basis for a training set for QSAR toxicity predictions. This dataset also enabled a detailed investigation of the correlations between chemical, functional, and structural properties of toxic compounds. We found that highly toxic compounds possess a higher molecular weight and more hydrogen bond donors and acceptors as compared to less toxic compounds, drugs, or natural compounds. Furthermore, an increased occurrence of certain functional groups and structural properties (e.g. chiral centers) was detected in highly toxic compounds. These “toxicity properties” form a very promising basis for the prediction of the toxicity of unknown compounds.

2. Data and Methods

2.1. Data

As no comprehensive collection of toxic compounds is publicly available, no sufficient information regarding the structural and physicochemical mutuality of toxic compounds is obtainable. Therefore, we collected more than 50,000 toxic compounds from literature and different web resources and stored them in a database. This provides a unique collection of data which enables an extensive investigation of their structural and physicochemical attributes.

The dataset was enlarged adding natural compounds and drugs from our databases [4,5]. Toxins often show similar modes of action to drugs which makes them ideal lead structures for the development of new drugs in pharmaceutical research. On the other hand, drugs as well as toxins are natural compounds or derivatives thereof. Since these substance classes are related, it seems appropriate to consider all of them in our investigations regarding similarities and differences.

Toxic compounds

To investigate the relation between a molecule's structure and its toxic impact, we compiled a database containing about 50,000 small molecule compounds, their structures and experimentally measured values of toxicity.

About 44,000 structures were taken from the Developmental Therapeutics Program (DTP) of the National Cancer Institute (NCI). Each compound was tested on 60 different cancer cell lines. Values for growth inhibition (GI50), total growth inhibition (TGI) and lethal concentration (LC50) were collected. Structures and toxicity information are freely available on the DTP website [6].

Toxicity information for about 4,500 molecules was extracted from the NLM [7] whereas corresponding molecular files were taken from PubChem [8].

Furthermore, about 1,200 structures were taken from the literature [9] and the corresponding toxicity values were extracted from the text.

The toxic compounds were investigated regarding their chemical and physicochemical properties. To elucidate the correlation of these properties, the compounds were subdivided into three groups according to their toxicity (-log (LC50)). Compounds with -log (LC50) values of 3 until 6 were combined in a slightly toxic group. The group of medium toxicity comprised compounds with -log (LC50) values of 6 until 9. The third group contained the highly toxic compounds integrating compounds which feature -log (LC50) values above 9. For a more detailed investigation the group of medium toxic compounds was subdivided into compounds with -log (LC50) value intervals between 6, 7, 8, and 9.

Drugs

For the comparative analyses we used the structures of about 2,500 drugs. These data were extracted from the free database SuperDrug containing WHO-classified drugs [5,10]. Entire plants, extracts, mixtures, colloids, and biopolymers are not included in this dataset.

Natural compounds

A second reference group is composed of natural compounds. About 47,000 structures were taken from the free database SuperNatural containing natural compounds, derivatives, and analogues [4].

This complete dataset enables the investigation of chemical, functional and structural properties and sheds light on the complex topic toxicity. Subsequently, the attributes of natural compounds and drugs will be discussed in relation to those of toxic compounds.

2.2. Methods

Calculation of chemical properties

The calculations of chemical properties, e.g. molecular weight, number of hydrogen bond donors and acceptors were performed with functions from OpenBabel 2.1, an open source chemical toolbox [11, 12]. To compute the properties for the structures the software MyChem was used which is an implementation of the OpenBabel 2.1 library for MySQL [13].

Analysis of functional and structural properties

For the analysis of functional and structural properties SMARTS patterns encoding functional and structural elements were defined [14]. The distributions of these patterns were analyzed between the different groups: the three groups of toxic compounds, the drugs, and the natural compounds.

3. Results and Discussion

3.1. Chemical Properties

Toxicity

In this study the toxicity of the compounds is defined as $-\log(\text{LC50})$, the medium deathly concentration for exposed animals or cells. The maximum values were reached in the $-\log(\text{LC50})$ class “4-5” (Figure 1), indicating that most of the compounds fell within the lower toxic range. As the distribution of 50,000 compounds is plotted, the numbers of those assembled in the classes 6-9 and higher still comprises of thousands of compounds.

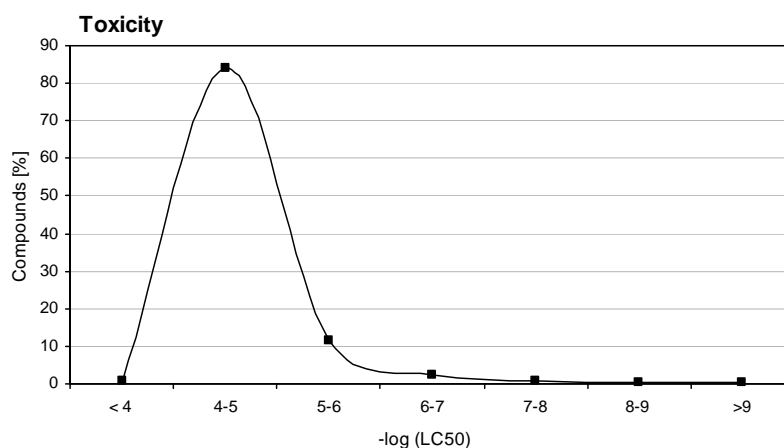


Figure 1. Distribution of the compounds according to their toxicity. The ratio of compounds is plotted against the $-\log(\text{LC50})$ values as a measurement of toxicity.

Molecular weight

Figure 2 depicts the distribution of the molecular weight of natural compounds, drugs and toxic compounds. It is noteworthy that the drugs have the lowest weight followed by the group of slightly, medium, and highly toxic compounds whereas natural compounds represent intermediate weights.

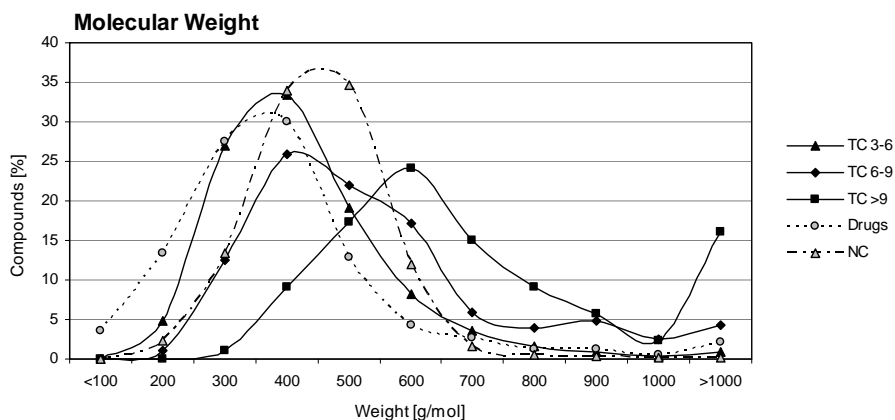


Figure 2. Distribution of the molecular weight of toxic compounds (TC), natural compounds (NC) and drugs. The toxic compounds are split into three classes according to their toxicity values ($-\log(\text{LC}_{50})$): 3-6 = slightly toxic, 6-9 = medium toxic, >9 = highly toxic).

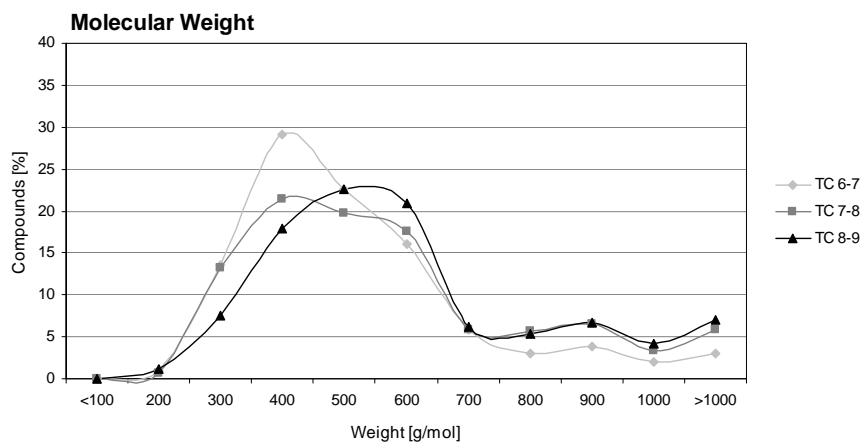


Figure 3. Detailed distribution of the molecular weight regarding the group of medium toxicity ($-\log(\text{LC}_{50})$: 6-9). (TC - toxic compounds)

Figure 3 shows a detailed distribution of the medium toxic compounds regarding their molecular weight. This diagram reflects the same trend as shown in Figure 2. The slightly toxic compounds are characterized by a lower molecular weight compared to the more toxic compounds. These findings support the tendency that toxic compounds have a higher molecular weight than non-toxic compounds.

In summary, the investigated groups of compounds differ according to their molecular weight forming a clear sequence: drugs, slightly toxic compounds, natural compounds, medium toxic compounds, and highly toxic compounds. Thus, a clear correlation between the toxicity and the molecular weight can be found. As drugs are designed as small molecules which can enter cells easily, these compounds are comparatively small. Within the highly toxic compounds, toxins like valinomycin (*Streptomyces fulvissimus*) or halichondrin (*Axinella sp.*) can be found. These large compounds function by binding to receptors or forming pores in membranes and are, therefore, very effective resulting in a high toxicity.

Hydrogen bond donors and acceptors

Hydrogen atoms attached to a relatively electronegative atom gain a positive partial charge which makes them very reactive. Thus, they act as hydrogen bond donors in the formation of a hydrogen bond to electronegative atoms such as fluorine, oxygen, or nitrogen which serve as hydrogen bond acceptors. Hydrogen bond donors and acceptors are ideal components of toxic compounds due to their high reactivity. Therefore, toxic compounds are even active at very low concentrations by interacting with biological macromolecules such as enzymes or cellular receptors.

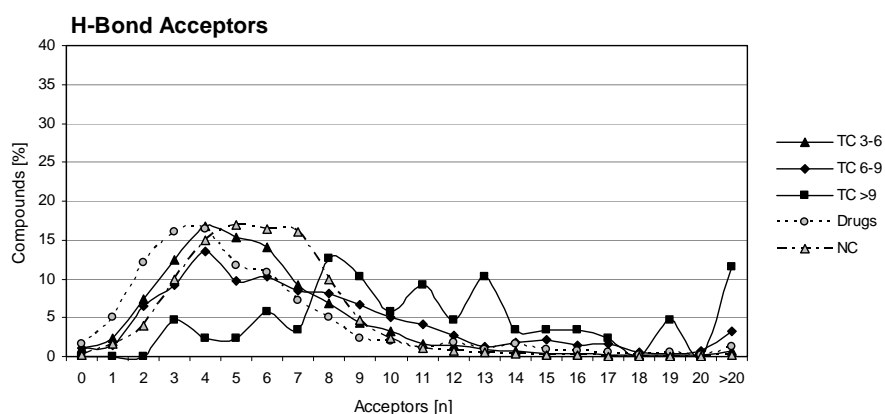


Figure 4. Distribution of the amounts of hydrogen bond acceptors of toxic compounds (TC), natural compounds (NC) and drugs. The toxic compounds are split into three classes according to their toxicity values (-log (LC50): 3-6 = slightly toxic, 6-9 = medium toxic, >9 = highly toxic).

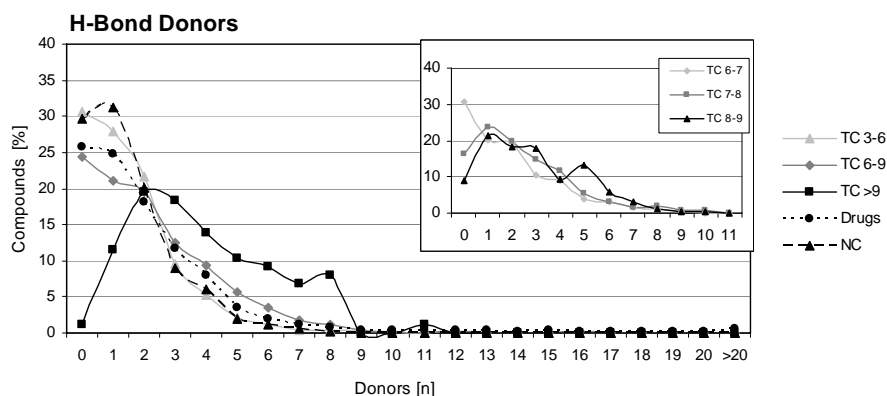


Figure 5. Distribution of the amounts of hydrogen bond donors of toxic compounds (TC), natural compounds (NC) and drugs. The toxic compounds are split into three classes according to their toxicity values ($-\log(\text{LC}_{50})$): 3-6 = slightly toxic, 6-9 = medium toxic, >9 = highly toxic). The small diagram shows a detailed distribution of the amounts of hydrogen bond donors regarding the group of medium toxicity ($-\log(\text{LC}_{50})$: 6-9).

To analyze this supposition, the amount of hydrogen bond donors and acceptors was compared between toxic compounds, natural compounds, and drugs (Figures 4 and 5). It was found that the group of natural compounds, slightly and medium toxic compounds, and drugs have very similar amounts of hydrogen bond acceptors as well as donors, ranging between three and six hydrogen bond acceptors and between zero and two hydrogen bond donors. The lowest number of hydrogen bond acceptors was found within drugs, as they are chemically designed to fulfill the Lipinski's rule of five [15]. According to this rule, they are supposed to comprise not more than 10 hydrogen bond acceptors in order to have adequate ADME properties [16]. In contrary to this, the group of highly toxic compounds shows both, more hydrogen bond donors and acceptors. It is obvious that within the groups of slightly, medium, and highly toxic compounds the amount of hydrogen bond acceptors and donors rises. This was confirmed by a more detailed investigation of the medium toxic compounds which show the same trend regarding the hydrogen bond acceptors (data not shown) and donors (Figure 5 small graph).

Comparing the molecular weight and the hydrogen bond acceptors the same sequence of compound groups can be found: the drugs feature the least amount of hydrogen bond acceptors followed by the slightly toxic compounds, natural compounds, and the medium toxic compounds concluding with the highly toxic compounds as the group with the highest amount of hydrogen bond acceptors. The same order occurs regarding the hydrogen bond donors, except that the natural compounds show the least amount of hydrogen bond donors and the drugs follow the slightly toxic compounds. Thus, the assumption was confirmed, that the more toxic a compound the more hydrogen bond donors and acceptors can be found in the structure.

3.2. Functional Properties

The distribution of functional groups in toxic compounds, drugs and natural compounds was analyzed and is depicted exemplarily in Figure 6. It can clearly be seen, that the occurrences of functional groups rises with increasing toxicity whereas the natural compounds and the drugs exhibit frequencies among those of the toxic compounds.

The highly toxic compounds differ significantly in the amounts of alcohol and sugar groups compared to the other compounds. The more hydroxyl groups can be found in a molecule, the more hydrogen bond donors are available and the higher is the reactivity.

Sugar molecules have many chiral centers and therefore, are characterized by a high stereo selectivity. Regarding the huge amount of different sugar molecules there is a vast number of possible combinations resulting in a high specificity according to the binding affinity to their targets.

Alcohol or phenol as an aromatic alcohol are characterized by their reactivity and corrosiveness resulting in a high toxicity. These properties are explained by the denaturing effect of phenol on membrane proteins forming pores which may lead to cell death.

Acetal includes a hydroxyl group which, as mentioned above, makes molecules more reactive. Acetals are stable with respect to hydrolysis by bases. This is an important property for toxic compounds since the more protected they are from hydrolysis the better they can perform their effects.

In summary, an order can be defined, starting with the slightly toxic compounds with the least amounts of the depicted functional groups followed by the natural compounds, the drugs, and the medium toxic compounds concluding with the highly toxic compounds which possess the highest frequencies of the mentioned functional groups.

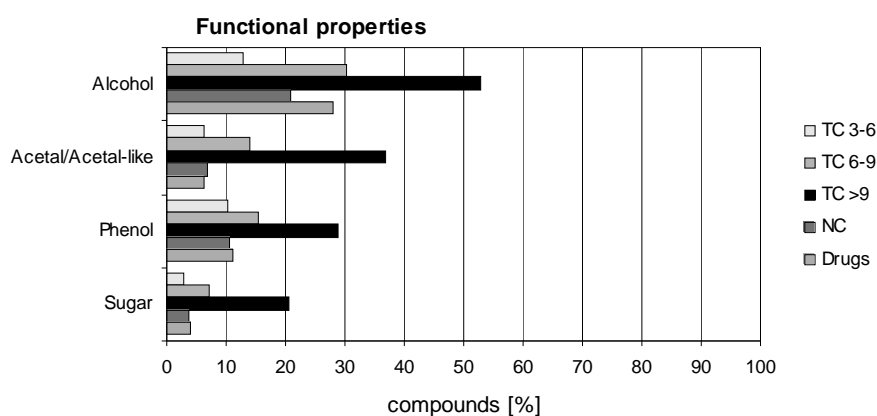


Figure 6. Distribution of the occurrences of functional groups of toxic compounds (TC), natural compounds (NC) and drugs. The toxic compounds are split into three classes according to their toxicity values (-log (LC50): 3-6 = slightly toxic, 6-9 = medium toxic, >9 = highly toxic).

3.3. Structural Properties

Structural properties were also investigated as toxicity indicators. The most distinct ones are represented in Figure 7. The analyses of the structural characteristics in the three groups of toxicity show results analogous to the analyses of the functional properties: the more toxic a compound the more distinctive the property.

Since chiral centers can be found in high amounts in sugar molecules their distributions correlate with those of the sugar group having the same origin: the high specificity and selectivity they provide ensure a very efficient and specific mode of action of toxic compounds.

Conjugated double bonds contribute to the stability of a molecule so that a high amount hampers degradation and enables the toxin to perform its effects.

Earlier studies revealed that the center of aromatic rings act as hydrogen bond acceptors [17] which is expected to play a significant role in molecular associations. This ensures a very specific and selective mode of action which explains the increasing amount of ring systems with increasing toxicity.

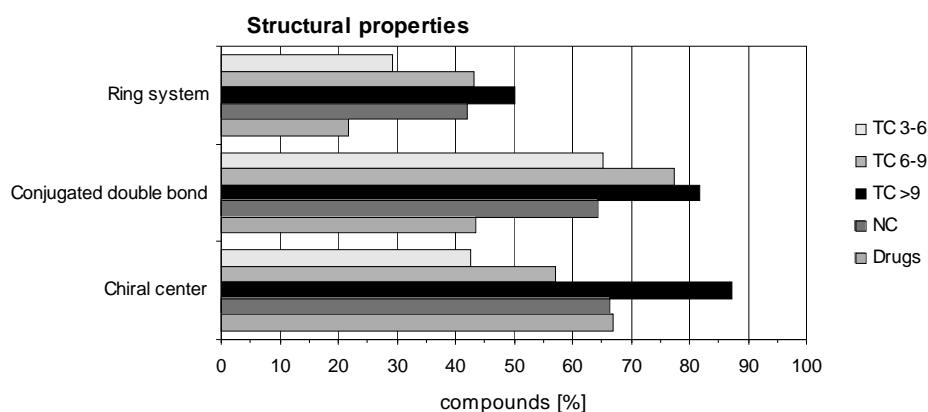


Figure 7. Distribution of the occurrences structural properties of toxic compounds (TC), natural compounds (NC) and drugs. The toxic compounds are split into three classes according to their toxicity values ($-\log(LC50)$): 3-6 = low, 6-9 = medium, >9 = high).

3.4. Case study

Amatoxins are cyclic non-ribosomal oligopeptides found in several members of the *Amanita* genus of mushrooms, one being the Death cap (*Amanita phalloides*). The most

deadly of all the amatoxins is the α -amanitin with an oral LD50 of approximately 0.1 mg/kg. It is an inhibitor of the RNA polymerase II blocking the transcription of DNA and RNA [18]. This leads to a total failure of the protein synthesis causing severe effects on liver and kidney [19]. Death usually occurs around a week from ingestion [20]. A map of the purine and pyrimidin pathway which can be found in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [21] is shown in Figure 8. It displays in detail the function of the RNA polymerase II and the effects its inhibition by α -amanitin would cause.

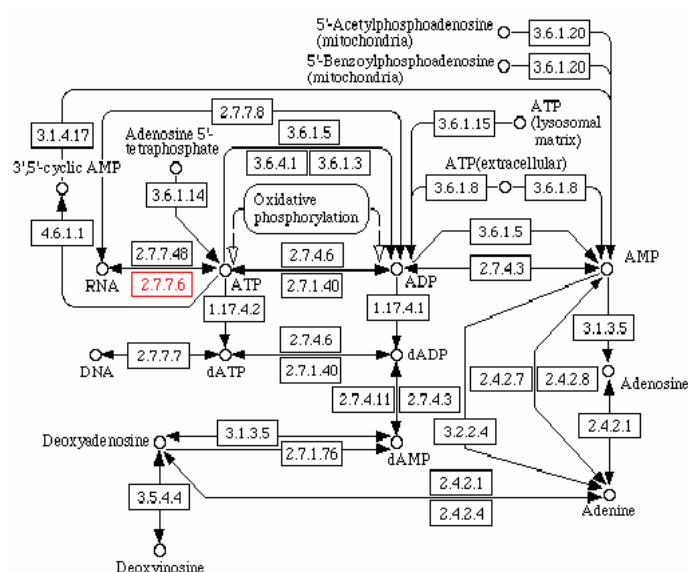


Figure 8. Excerpt of the purine pathway extracted from KEGG. The enzyme colored in red with the number “2.7.7.6” depicts the RNA polymerase II.

With a molecular weight of 918.97 g/mol, 13 hydrogen bond donors, and 15 hydrogen bond acceptors the chemical “toxicity properties” of α -amanitin are consistent with our findings of the highly toxic compounds. A lot of ring systems, conjugated double bonds, and chiral centers also fit in our results of the structural “toxicity properties” of the highly toxic compounds.

4. Conclusion and future perspectives

In this work we were able to elucidate a continuous trend in structural, chemical, and functional properties within the different groups of toxic. The analysis of hydrogen bond donors and acceptors as well as certain functional groups and structural features revealed

a positive correlation between occurrence and toxicity whereas the amounts of drugs and natural compounds have similar values compared to the slightly toxic compounds.

Toxic compounds function in a variety of ways and subgroups, like the highly toxic ones, react with their target in a completely different manner than drugs. While drugs are usually small compounds, able to enter the cell and to affect targets within the cells, a lot of toxic compounds function by forming pores in membranes (e.g. alpha toxin from *Staphylococcus aureus*), by permanent activation of for example sodium channels (aconitin) or by interaction with neurotransmitter receptors (strychnin). With the help of such mechanisms these toxic compounds are able to affect critical pathways which often cannot be circumvented. Therefore, these molecules are very effective.

The data presented here provide valuable insight into the phenomenon of toxicity by elucidating “toxicity properties”, characteristics of toxic compounds. Thus, the properties analyzed here will function as additional criteria to predict toxicities with the help of QSAR. Additional toxicity relevant properties, as presented here, will be helpful to improve such analysis. Further efforts will be made in the prediction of potential targets of unknown compounds.

Acknowledgements

This work was supported by the International Research Training Group Boston-Kyoto-Berlin, funded by the German Research Foundation (DFG).

References

- [1] Watson, P., Spooner R.A., Toxin entry and trafficking in mammalian cells, *Adv Drug Deliv Rev*, 58: 1581-1596, 2006.
- [2] Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., Su, Z., Perkins and R., Tong, W., Mold(2), Molecular Descriptors from 2D Structures for Chemoinformatics and Toxicoinformatics, *J Chem Inf Model*, 2008.
- [3] Hughes, L.D., Palmer, D.S., Nigsch, F. and Mitchell, J.B., Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P, *J Chem Inf Model*, 48: 220-232, 2008.
- [4] Dunkel, M., Fullbeck, M., Neumann, S. and Preissner, R., SuperNatural: a searchable database of available natural compounds, *Nucleic Acids Res*, 34: D678-683, 2006.
- [5] Goede, A., Dunkel, M., Mester, N., Frommel, C. and Preissner R., SuperDrug: a conformational drug database, *Bioinformatics*, 21: 1751-1753, 2005.
- [6] <http://dtp.nci.nih.gov/>
- [7] <http://chem.sis.nlm.nih.gov/chemidplus>
- [8] <http://pubchem.ncbi.nlm.nih.gov/>
- [9] Teuscher, E. and Lindequist, U., Biogene Gifte, Gustav Fischer Verlag, Germany, 1994
- [10] Gunther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, J., Urdiales, E.G., Gewiess, A., Jensen, L.J. *et al.*, SuperTarget and Matador: resources for exploring drug-target relationships, *Nucleic Acids Res*, 36: D919-922, 2008.

- [11] Guha, R., Howard, M.T., Hutchison, G.R., Murray-Rust, P., Rzepa, H., Steinbeck, C., Wegner, J. and Willighagen, E.L., The Blue Obelisk-interoperability in chemical informatics, *J Chem Inf Model*, 46: 991-998, 2006.
- [12] <http://openbabel.sourceforge.net/>
- [13] <http://mychem.sourceforge.net/>
- [14] <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>
- [15] Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv Drug Deliv Rev*, 46: 3-26, 2001.
- [16] van de Waterbeemd, H. and Gifford, E., ADMET in silico modelling: towards prediction paradise?, *Nat Rev Drug Discov*, 2: 192-204, 2003.
- [17] Levitt, M. and Perutz, M.F., Aromatic rings act as hydrogen bond acceptors, *J Mol Biol*, 201: 751-754, 1988.
- [18] Lindell, T.J. et al., Specific Inhibition of Nuclear RNA Polymerase II by agr-Amanitin, *Science*, 170: 447-449, 1970.
- [19] Wieland, T., Poisonous Principles of Mushrooms of the Genus Amanita: Four-carbon amines acting on the central nervous system and cell-destroying cyclic peptides are produced, *Science*, 159: 946-952, 1968.
- [20] Mas, A., Mushrooms, amatoxins and the liver, *Journal of Hepatology*, 42: 166-169, 2005.
- [21] <http://www.genome.jp/kegg/>