# ASSESSMENT OF SOLUBILITY AS A PHYSICOCHEMICAL CLASSIFYING PARAMETER IN THE ANATOMICAL THERAPEUTIC CLASSIFICATION

Francesca Pace, Frederick Lia and Claude Farrugia
Department of Chemistry, University of Malta, Msida MSD 2080, Malta

## Introduction

The Anatomical Therapeutic Chemical (ATC) classification system managed by the World Health Organisation is nowadays considered the most widely used drug classification system. Medicinal compounds are classified into different groups according to the organ or system on which they exert their effect on, also taking into consideration the therapeutic, pharmacological and chemical property of the active ingredient (Chen *et al.*, 2012; WHO, 2015). The objective of the study was to extend the studies performed to date by Fenech and Farrugia (2014) to all of the medicinal compounds found in the ATC Classification to assess whether drug solubility can be included in common physicochemical profiles existing for compounds classified within the same ATC sub-group.

## Methodology

A structured approach was taken so as to explore the correlation between the physicochemical parameters of drugs classified within the same ATC sub-group. The medicinal compounds used in this study were first identified. For each identified medicinal substance, seven physicochemical parameters, namely, molecular density, total surface area (TSA), polar surface area (PSA), Log P, parachor, molecular weight (MW) and solubility at pH 7.4, were generated using a computational approach. Density and parachor were generated using ChemSketch (*Advanced Chemistry Development Inc., ACD/Labs*); total surface area (TSA), polar surface area (PSA), molecular weight (MW) and Log P were obtained using *Chemicalize.org by ChemAxon* whilst solubility at pH 7.4 was generated by using MarvinSketch (*ChemAxon)*. The generated physicochemical data was then statistically evaluated using several statistical methods, including Multivariate Platform, Principal Component Analysis (PCA) and K-Means Clustering using JMP software. The correlation between the medicinal compounds found in the ATC level 4 sub-group and their physicochemical parameters was statistically computed using JMP software.

## Results and Discussion

### Multivariate Platform

Using multivariate platform analysis and a scatterplot matrix, a positive linear correlation was found between TSA, PSA, parachor and MW, while an inverse correlation was exhibited between Log P and density, and Log P and PSA. No significant correlation was exhibited between solubility and any other property.

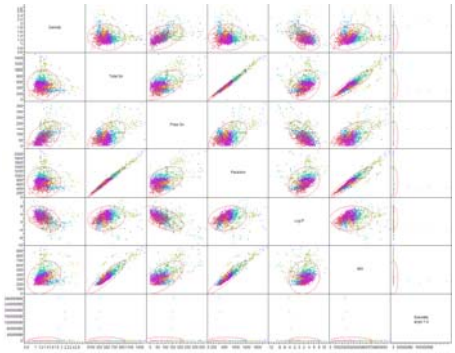


**Figure 1: A scatterplot matrix explaining visually the correlation between the physicochemical properties of the medicinal substances.**

| | Density | Total SA | Polar SA | Parachor | Log P | MW | Solubility at pH 7.4 |
|---|---|---|---|---|---|---|---|
| Density | 1.0000 | -0.0724 | 0.4877 | -0.0290 | -0.4075 | 0.2636 | 0.1066 |
| Total SA | -0.0724 | 1.0000 | 0.5169 | 0.9883 | 0.1929 | 0.9203 | -0.0390 |
| Polar SA | 0.4877 | 0.5169 | 1.0000 | 0.5171 | -0.5876 | 0.6216 | 0.0547 |
| Parachor | -0.0290 | 0.9883 | 0.5171 | 1.0000 | 0.2157 | 0.9456 | -0.0417 |
| Log P | -0.4075 | 0.1929 | -0.5876 | 0.2157 | 1.0000 | 0.1099 | -0.0971 |
| MW | 0.2636 | 0.9203 | 0.6216 | 0.9456 | 0.1099 | 1.0000 | -0.0283 |
| Solubility at pH 7.4 | 0.1066 | -0.0390 | 0.0547 | -0.0417 | -0.0971 | -0.0283 | 1.0000 |

**Table 1: A multivariate model correlation table explaining the correlation between the physicochemical properties of the medicinal substances.**

### Principal Component Analysis

A PCA of all variables gave a first principal component that accounted for 75% of the total variance, and was characterized by major positive levels of MW, parachor, PSA and TSA. In the second principal component, Log P exhibited a high negative value while density showed a high positive load. Density was highly loaded in the third principal component, but solubility did not exhibit a significant loading in any of the components. These findings were further confirmed through a Rotated Factor Loading matrix.

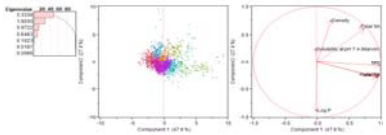


**Figure 2: Eigenvalues, Score Plot and Factor Plot obtained for all of the medicinal compounds classified according to their physicochemical properties.**

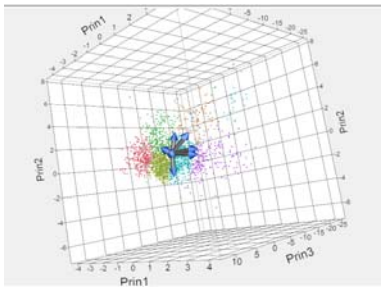


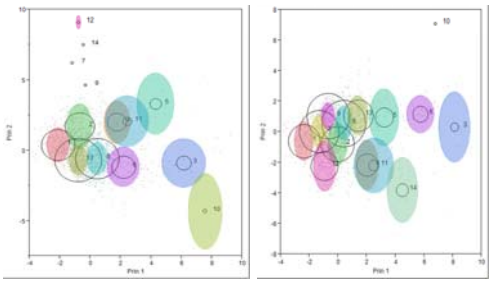

**Figure 3: A Score Plot with three Principal Component Analysis obtained for all the drugs classified according to their physicochemical properties.**

| Rotated Factor Loading | | | |
|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 |
| Density | 0.079248 | -0.244033 | 0.953936 |
| Total SA | 0.978835 | 0.038411 | -0.147459 |
| Polar SA | 0.597776 | -0.662214 | 0.292439 |
| Parachor | 0.992641 | 0.073639 | -0.093983 |
| Log P | 0.130175 | 0.914728 | -0.203977 |
| MW | 0.969746 | 0.027172 | 0.202738 |
| Solubility at pH 7.4 (Marvin) | -0.028098 | -0.082815 | 0.084992 |

**Table 2: A Rotated Factor Loading table explaining the values of each physicochemical parameter for different factor analysis.**

### K-Means Clustering

In a K-Means Clustering analysis it is seen from Figures 4 and 5 that when solubility is included more are compounds classified as outliers; these were compounds with very high value of solubility. When solubility was excluded, only one drug was an outlier. The cluster analysis was also extended to excluding each parameter individually so as to distinguish which parameter which impacts the cluster analysis negatively. From the Eigenvalues obtained, it could be seen that the exclusion of solubility resulted in a sharp increase in the percentage variation explained with the second component, to levels that could only be achieved with the third component in other cases.



**Figures 4 and 5: K-Means Clustering for all the medicinal compounds with all their physicochemical parameters and with all the physicochemical properties excluding solubility at pH 7.4 respectively.**

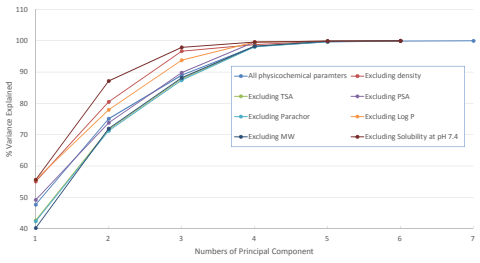



**Figure 6: Change in percentage variance explained by principal components for models with a single physicochemical variable excluded.**

## Conclusion

From the evaluation of the different computational analysis, it can be seen that solubility at pH 7.4 has no correlation with any other physicochemical parameter. Hence, the above-mentioned parameter should be excluded from further studies to analysing the correlation between the physicochemical parameters of the medicinal compounds to their Anatomical Therapeutic Chemical Classification.

## References


- Fenech, M.; Farrugia, C.A (2014). M.Sc. Dissertation, Department of Chemistry, University of Malta, Malta.
- Chen, L.; Zeng, W.; Cai, Y.; Feng, K.; Chou, K (2012). PLoS One, 7(4), 1-7.
- Chemicalize.org by ChemAxon (http://www.chemicalize.org/)
- ACD/ChemSketch (Freeware) 2012, version 14.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2013
- MarvinSketch 15.2.2.0, 2015, ChemAxon (http://chemaxon.com)
- JMP® 10.0.0, 2012, SAS Institute Inc., Cary, NC, 1989-2007