Applied method of extending van Krevelen diagrams for exhaustive analysis of metabolites

Kaori Matsuda¹
kaori-m@is.naist.jp

Md.Altaf-Ul-Amin¹
amin-m@is.naist.jp

Kensuke Nakamura¹
kensuke-mm@is.naist.jp

Hiroki Takahashi¹
hi-takah@bs.naist.jp

Daisaku Ohta²
ohtad2g30490@bioinfo.osakafu-u.ac.jp

Shigehiko Kanaya¹
skanaya@gtc.naist.jp

¹ Graduate School of Information Science, NAIST, 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan
² Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai 599-8531, Japan

Keywords: metabolite, classification of metabolic and enzyme response, chemoinformatics

1 Introduction

The exhaustive analysis of metabolites in the context of metabolic pathways is very important. Thus, the objective of our work is to make possible visual comparison of reaction paths of various metabolites by extended van Krevelen diagrams. The van Krevelen diagram is a two dimensional (2D) scatter plot which has axes on the basis of the ratio of the number of major elements (e.g. H/C, O/C etc.) in molecules. According to earlier studies, it is useful to arrange the metabolic paths and to classify features of the structure for the identification of metabolites. So, we try to comprehensively understand features of metabolites taking a type of functional group, a type of reaction, chemical species, etc, into consideration by extended van Krevelen diagrams. In the present study, we try to clarify difference between atomic composition of molecules belonging to natural and synthesized products based on extended van Krevelen diagrams. Furthermore we extend the original van Krevelen diagrams visualized by 2-dimensional space to multi-dimensional space by considering the elements other than H, C, and O. We analyzed sets of metabolites accumulated in databases such as KNAPSAcK and KEGG. Using it, we investigate the coordinate patterns concerning the structure and species of metabolites, and vector patterns concerning the reaction of metabolites and enzymes.

2 Method and Results

Analysis by the van Krevelen Plots needs chemical formula of metabolites. In the present work we used four chemical structure databases namely KNAPSAcK, KEGG drug, KEGG compound, and KEGG reaction each of which has different characteristics. KNAPSAcK has secondary metabolites of plants, KEGG drug has approved drugs in Japan and America, KEGG compound has metabolic substances in biological systems and KEGG reaction has substance pairs (reactant and product) corresponding to chemical reactions in biological systems.

2.1 Method

We try to differentiate coordinate ranges concerning several elements contained in the metabolites on a van Krevelen Plot by extending it to 3D for certain purposes. (Table. 1) It is useful to analyze how chemical reactions affect changes of structures or addition/removal of functional groups etc.

Table 1: Axes for 3D van Krevelen Plots

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O/C</td>
<td>H/C</td>
<td>(a)/C (a) = elements are a lot in composition formula</td>
</tr>
<tr>
<td>2</td>
<td>(b)/C</td>
<td>H/C</td>
<td>(b)/H (b) = elements of functional groups that are a lot in metabolites</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(for example, hydroxyl -O, amino –N, thiol –S, imino =N, oxid =O)</td>
</tr>
<tr>
<td>3</td>
<td>(c)/C</td>
<td>H/C</td>
<td>(c)/O (c) = elements of functional groups that are a lot in metabolites</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(for example, nitro –N, sulpho -S, phosphate -P)</td>
</tr>
</tbody>
</table>
2.2 Results

We try to analyze the relationship between coordinates and elements on van Krevelen plots. Many molecules in the dataset used in this work contain Nitrogen(N), Sulfur(S), Phosphorus(P) and Chlorine(Cl), other than Carbon(C), Hydrogen(H) and Oxygen(O). Distribution of molecules on coordinate space containing different elements are different as shown in Fig. 1, which indicates when any other element is added to functional group or hetero ring then that influences O/C and H/C ratios.

Using extended van Krevelen plot, accommodates almost all substances in narrow coordinate space. (Fig. 2(a)) and ratios corresponding to each molecule is related by simple mathematical expression.

\[ Z(\frac{b}{H}) = a[\text{coefficient}] \times X(\frac{b}{C}) \]

Fig. 2(b) shows chemical binding rule have linearity. It indicates that classification by coordinate space is possible in spite of variation in binding elements.

3 Discussions

This work indicates that rules of chemical bonding are describable by simple mathematical expression. Thus, in the future work, we try to consider the distribution pattern of molecules in coordinate space for supervised learning because of clear relationship between coordinate distribution and molecule types and also characteristic coefficients of mathematical expression depend on molecules. Moreover, we try to extract characteristic features from well-known data and in the present work we try to determine those by using data from KNApSAcK, KEGG, PubChem databases. These findings might be important to classify unknown data.

References

