Massila Kamalrudin, Soong Hoong Cheng,
Azlianor Abdul Aziz, Muhammad Suhaizan Sulong

Reinforcing the Concept of Calculating Isotope Pattern Using Theoretical Isotope Generator (TIG)

MASSILA KAMALRUDIN
SOONG HOONG CHENG
AZLIANOR ABDUL AZIZ
MUHAMMAD SUHAIZAN SULONG
Faculty of Information and Communication Technology (FICT)
Universiti Teknikal Malaysia Melaka (UTeM)
Locked Bag 1200, Hang Tuah Jaya, Ayer Keroh, 75450, Melaka
MALAYSIA
massila@utem.edu.my http://www.utem.edu.my/

Abstract: - This paper describes how the isotope patterns are calculated for chemists, students, lecturers and researchers using Theoretical Isotope Generator (TIG) integrated from the .NET framework (specifically VB.NET) with the data from the dual databases (Microsoft Access and Microsoft SQL 2000). Accordingly, TIG in terms of its functionality and features is an application developed to gain related information on molecules and their relative intensity for educational purposes (chemists, students, lecturers and researchers). Although various mass spectrometer applications are available as web-based and windows-based systems, the projected software is indeed practical, beneficial and constructive for the academic learners as well for the intellectual researchers for rapidly defining the numbers of the possibility isotope patterns obtainable from a substance(chemical compound) for example methane (CH_2). Therefore, TIG with the mathematical application called Cartesian Products is used to obtain the isotope patterns of any atom or molecules. TIG provides extra features than others that are proficient to retrieve the information of the mass (atomic/molecular weight) from the substance specified based from the isotope patterns available. Additionally, TIG is completely with several functionalities such as drawing, normalizing and zooming the generated graph that convey with the molecular information in a number of formats by providing the details of the calculation and molecules.

Key-Words: - Isotope pattern, Cartesian Products, Calculation, Engine, Isotopic Distribution and Theoretical Isotope Generator (TIG).

1 Introduction

Indeed with the utterance of isotope, mass spectrum and mass spectrometer are associated firmly with one another. For the past existence years, mass spectrometer offers the defined capability to distinctly identify the convoluted atomic isotopes from the mass spectrum projected onto the screen. Conversely, even though the mass spectrometer does not necessitate to be implemented to the development software but it is crucial to be comprehended about the elementary ideas that are bonded indirectly of the aforementioned tool.

According to the definition by *The Cosmic Lexicon*, the **mass spectrum** confers the meaning as the pattern of relative's abundances of ions for different atomic or molecular mass (mass-to-charge ratio) within a sample and frequently refers to given element measured with the relative abundance of isotopes. The source also point out that a mass spectrum (or m/z) on the x-axis and corresponding ion abundance on the y-axis as a series of peaks [1].

Therefore, the peak is corresponding to the ion greatest abundance is called based peak which can correspond to the molecular ion or to any one of the fragment ion. Essentially, the mass spectrum is derived from mass spectrometer as it is shown from the illustration (Fig. 1).
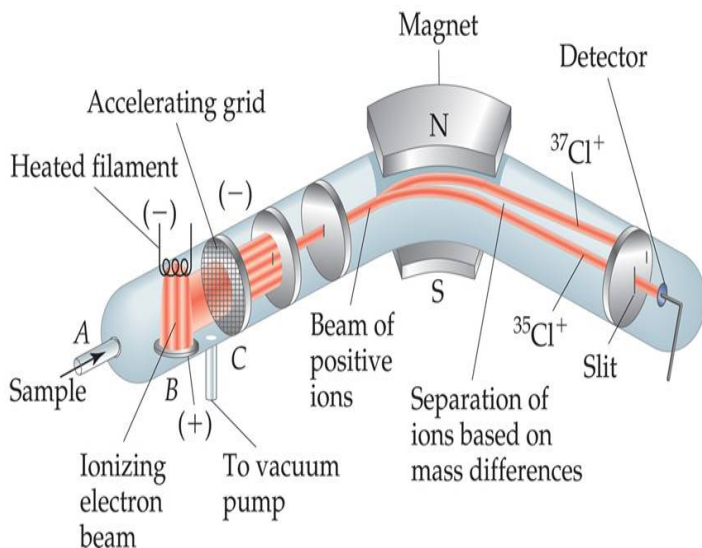
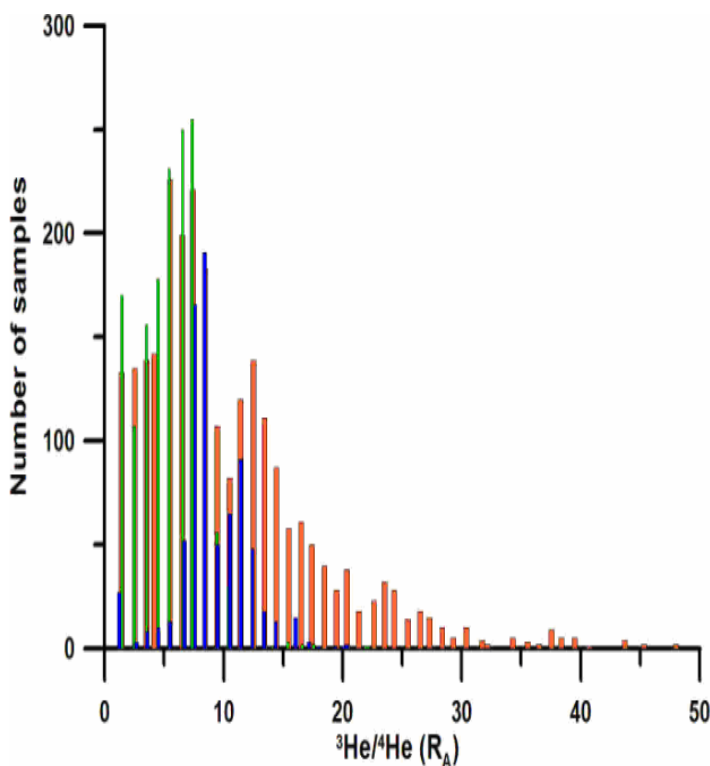Fig. 1 Mapping schematics of a simple mass spectrometer

Fig. 2 Mapping mass spectrum of the isotopic distribution

As for this point of view, based from *Ruedi Aebersold* and *Matthias Mann*, mass spectrometric measurements are carried out in the gas phase on ionized analytes. By definition, a **mass spectrometer**

consists of an ion source, a mass analyzer that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value. Hence, it is the composition of the physical sample (molecular substances) by creating a mass spectrum corresponding to the masses of the components [2]. Consequently, mass spectrometry is the most accurate and direct method for determining atomic and molecular masses (molecular weight) [18].

Based from the physics explanation, any of two or more atoms of an element are having the same number of protons, but a different number of neutrons in its nucleus. As a consequence, it will have the same atomic number but a different mass number (atomic weight). **Isotopes** are any of the several different forms of an element each having different atomic mass (mass number). Isotopes of an element have nuclei with the same number of protons (the same atomic number) but different numbers of neutrons. Therefore, isotopes have different mass numbers, which give the total number of nucleons; the numbers of protons plus neutrons. As for this, stable isotopes are chemical isotopes that are not radioactive. Accordingly, stable isotopes of the same element have the same chemical characteristics and therefore behave almost identically [3].

From mass spectrum the **Isotopic Distribution** (Fig. 2) is certainly to be expected from the graph subsequent to the devices that specifically yield the results in the form of graphs; as in y-axis is the signal intensity/ relative abundance and x is the mass for the substances. Hence, the peaks of the aforementioned graph accentuate the imperative masses representing the isotopic elements from the periodic table. As a result, the mass spectrometer undeniably attests to all for examining all the possible isotopes in the atomic or molecular substances. Typically, the highest abundance of the peak is the epitome of the usual atomic/molecular substances whereas the other lower peaks are the isotopic subsisted substances from the particular elements. As each peak is determined, it is essential scrupulously to compare the masses to avoid any misinterpreting elements from the sample substances.

Isotopic distribution allows predicting the masses and abundances of the isotopic formula. Hence, mass spectrum helps in comparing data from mass spectrometer and to predict the number of atom of a specific element in the formula that contributes to the combination of molecules [17].

2   Related Research

Currently, there are indisputably numerous systems that are obtainable in both web-based and windows-based system with the interactive isotope calculators as the user interfaces. With the inferences from the existing systems such as Sheffield **ChemPuter, Isotope Distribution Calculator, PAT for windows and Isotope Pattern Calculator v4.0, the mass spectrometer** is not implicated with the mentioned systems yet the theoretical of the isotopic distribution is visibly projected from the mass spectrometer and implied. To be concise, the related research concerns about isotopic pattern calculators accessible for educational purposes [4].

As for the web-based **Sheffield ChemPuter**, it is a set of simple interactive calculators for chemistry, including isotope patterns, element patterns, reaction yields, oxidation states of d-block metals in a complex, electron accountancy, and more. Nevertheless, some modules are still in testing phase. Furthermore another windows-based isotope calculator system has been initiated to advocate a solution for the mass spectrum pattern calculator [6], thus **PAT** for windows is developed to be applicable in the research in order to identify the fragment ions of chlorocarbon which is derivative according to their chlorine isotope pattern in low mass spectra. Additionally, it has been utilized in the work of polynuclear metal compound that is to find the number of selenium atoms [7].

Parallel from the other identified isotope pattern calculators, PAT computes the pattern of intensities which will be produced as a line graph with a list of relative intensities against the mass number. Following of the creation, research purposes are not the foremost reason as it is pertinent to be as a learning tool for the related chemistry classes. For this reason, there are few steps to employ PAT. It is evaluated that PAT only contains incomplete set of molecules and it presents in a stack of five. So, user needs to use drop-down list box to show formula which windows are currently presented. Then, if you try to generate more than 20 windows, the earliest drop down list will be deleted. Although PAT has few advantages, PAT still unable aim for the full features as limitation occurs for the system [7].

Hence, **Isotope Pattern Calculator** (IPC) is a program that calculates the isotopic distribution of a given chemical formula. It gives the relative intensities and the probability of the masses belonging to a molecule ion, fragment or whatever is represented by the given chemical formula. Furthermore it can use GNU Plot to visualize the result. Only masses with a Relative Intensity bigger then 0.009% are shown. Additionally, IPC (Fig. 3) prints the overall number of peaks and the needed computation time. The program uses an algorithm which computes the exact isotopic distribution. This leads to a large number of peaks which have very low relative abundances. Even for a small molecule as Acetylsalicylic acid ($C_9H_8O_4$, Mass Relative=180.15) there are 1350 peaks but only nine of them have a relative abundance higher then 0.01% [5].
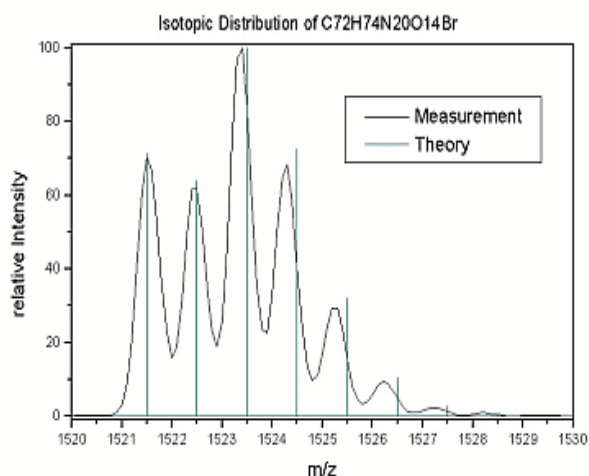

Fig. 3 Peaks of isotopic distribution generated from IPC

Alternatively, when using the Isotope **Pattern Calculator** the user initially needs to enter the molecules. At that moment, the results will come out with the most intense line that is set to 100%. Accordingly, the limitations of this sub function are the incidence of case sensitive when inputting molecules and the compound names as well as elements names for the formula. For instances, the element names such as water or iron is not acceptable prior to be inputted for the multifarious calculations [8].

From the observation of the process and output for the both calculators, the outcome is analogous as so for those other calculators offered in the present day. Sensibly to acquire the peaks of the isotopic distribution, those calculators tend to be requesting the molecule structures from the specified substances. Therefore, with the attainment from the

molecules structures the possibility patterns of the isotopic patterns are meticulously analyzed and defined. With the mathematical algorithms and chemistry knowledge applied to programs, the isotopic distribution is subsequently estimated onto the graphs.

As for the aforementioned calculators, these calculator systems are signifying assertion of the availability 24 hours per week at anytime or anywhere and hitherto the features are still limited to the capability of the browsers. And yet the disadvantages for the calculators such as PAT and Isotope Pattern Viewer, only the windows-based operating systems operate the calculator systems whereas Linux-based application is unquestionably inapplicable.

And so, these systems can be perceived to be further complex than the web-based calculators that provide more features as it is not dependable to the browsers. In consequence, all these systems are contrasted from each other and analyzed prior to suggest any further readings for the research.

There are also several algorithms are found to be applicable to the isotope calculator engine (**Z-score Algorithm, Polynomial Algorithm, Binomial Algorithm and Naïve Bayes Algorithm**):

1) Z-score Algorithm

According to *Zhongqi Zhang* and *Alan G. Marshall*, apart from the other algorithm the general Z-score algorithm would be finding the highest-abundance unprocessed spectral peak (h) that determine its maximum possible charge of state [9].

Then, for each possible charge state (z) assign a score by a particular charge scoring system. As the assign highest-scoring charge state is correct, find the isotope multiples of h that transform each peak in multiple to *zero-charge* mass domain and add each amplitude to prior amplitude at the same mass. As it is done, denote transformed peak(s) as "processed" and remove from original spectrum. The present method is quite general in that the user may adapt a particular scoring (weighting) system to suit a particular situation [9].

2) Polynomial Algorithm

Indeed, the polynomial method can be also applied for the small molecules such as water. The peaks are generated in the form of list and each element of the peak list is linked with the specific isotope composition and probability. The accurate mass of nominal isotopic peak would be calculated by sorting through the list to organize the list into isobars and then computing the mass of a nominal isotopic peak as a probability-weighted sum of all the terms within an isobaric group [10].

Although the polynomial method works well for small molecules, the recent trend in mass spectrometry has been toward larger and larger systems. For these, the polynomial method becomes increasingly complex and impractical [10].

3) Binomial Algorithm

Based from *Alan G. Marshall et al.*, natural abundance of ^{13}C is 1.066-1.106% relative to ^{12}C as 100% as for a molecule containing n carbons, the isotopic distribution is a binomial expansion (0.9889 + 0.0111)n, and it is ~n% as likely that a given molecule will contain one ^{13}C as that all of the carbons will be ^{12}C [11].

For protein with hundreds of carbon, hydrogen, nitrogen, and oxygen, the combined binomial distributions for its constituent elements produce a wide spread in natural-abundance isotopic relative abundances [11].

4) Naïve Bayes Algorithm

Referring to *Sean Mcllwain et al.*, the model is able to calculate the isotopic distribution probabilities too. The features employed by the model encode data about charge state, distance between peaks within a potential distribution, distance between the first peaks of the potential distribution and the last peak before it, distance between the last peak of the potential distribution and the next peak after it, the shape of (sequence of peak heights in) the potential distribution, the number of peaks in the potential distribution and the relative ratio of intensities of the two highest peak in the distribution (repeated also for the two lowest peaks) [12].

The remainder of this section describes these features in more details. In that case, Naïve Bayes model assumes these features are independent of one another given the class (true isotopic distribution or not an isotopic distribution). Even when this assumption is violated, Naïve Bayes works better than the complicated Bayesian Models in comparison [12].

3   Methods

From the theoretical aspects, the initial approach for the development is specifically studied for searching the suitable algorithm to be implemented into the system as the calculation method (*Cartesian Product Algorithm*). As it is discussed in detail in the calculation technique section, briefly from the calculation method as it calculates the possibility peaks that are generated from the isotopic distribution.

With the effectively implemented of the system, then the algorithm is implemented by using the programming language called VB.NET. According to *Nell Dale et al.*, **Visual Basic.NET** (VB.NET) is an object-oriented computer language that can be viewed as an evolution of Microsoft's Visual Basic (VB) implemented on the Microsoft .NET framework [15]. Hence, from the primary selection of the programming language that means half the specifications are automatically selected for window-based compliance applications. Even though, JAVA is more flexible to cross-platform the selection of the VB.NET affects the databases selection that are Microsoft Access and Microsoft SQL 2000 server.

For the motive for dual implementation of the databases, the **Microsoft SQL 2000** is more solely used for the administrative purposes as when to update information of the new atomic/molecular elements into the database for calculation purposes. Alternatively, the **Microsoft Access** is to ease users to utilize the database compared to the complex configuration of Microsoft SQL 2000.

3.1   Proposed system: System Framework / Architecture

Referring to the illustration above (Fig. 4), the system architecture consists of three major components that are users (lecturers, student, researchers or others), calculator system and lastly for the processed result. Within the calculator system, the calculator can be divided into two sub-components that are calculation engine utilizing Cartesian Product Algorithm and the dual databases are managed by the selected two database management system (DBMS). From the visible output as the result, the existences of the three projected outputs are normalized graph, normal graphs and details of the calculations.
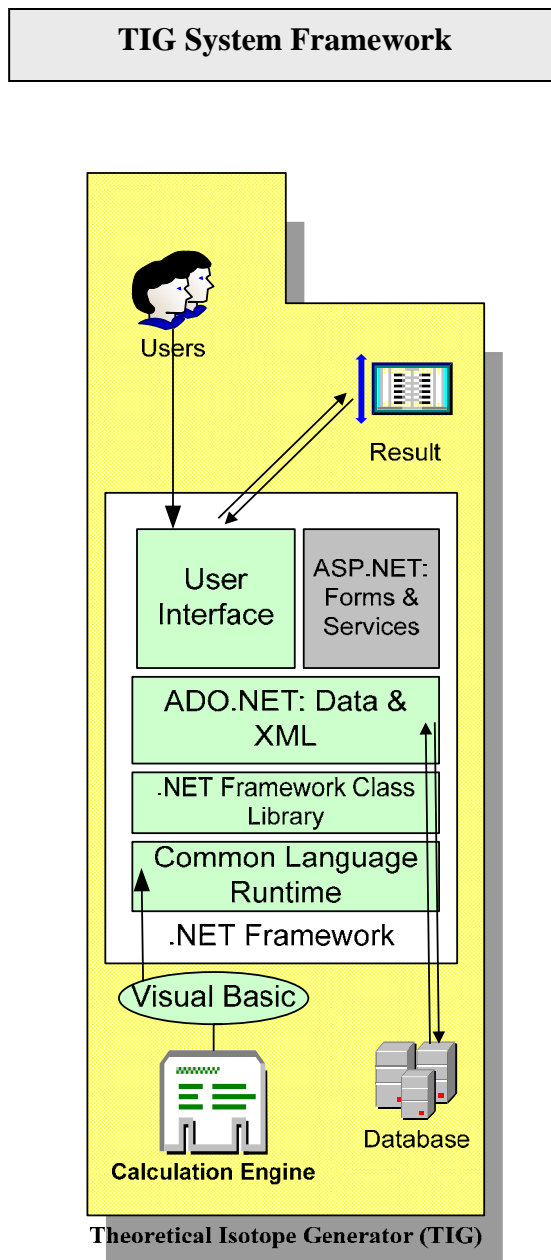

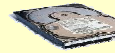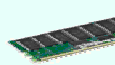Fig. 4 System architecture for Theoretical Isotope Calculator

According to Microsoft Official Course, the .NET Framework provides all the common services required to run your Visual Basic .NET applications. Since Visual Basic is chosen as the common language specification due to able to meet several goals (based on Web standards and practice, extensible, easy for developers to use and designed using unified application models) [16], the calculation engine (Fig. 4) is executed by Visual

Basic as the common language runtime for the .NET Framework. Therefore, with the data acquired for the calculations from user interface (input by user) and ADO.NET (data from database) the results are projected from the subsequent calculated process from the Theoretical Isotope Generator (TIG).

However, even if (ASP.NET) Forms and Services are from the .NET Framework may not be applied due to Visual Basic is chosen as the programming language, but if the Theoretical Isotope Generator is needed as the web-based application then the mentioned components will be used together with Visual Basic as the common language runtime from the .Net Framework.

As for the minimum specification for the hardware to run the system, it is listed orderly as the table below:

TABLE I
HARDWARE MINIMUM SPECIFICATION

Hardware Requirement	Details
Operating System	Windows 2000/NT/XP/VISTA
Processor	Pentium III or Above
Hard Disk	10.5 GB or greater
RAM	128 MB (Minimum requirement for Windows 2000)
Monitor	SVGA Monitor

3.2 Calculation Technique Used (Cartesian Product)

Due to the advancement of mathematics nowadays, each of the existing isotope calculators consists of at least one of the algorithm to calculate the isotope patterns from the specified substances. Despite the results may be similar, the performance and the rapidity of the processed output data in fact might vary depending from the equivalent system specifications.

$$X_1 \times \cdots \times X_n = \{(x_1, \ldots, x_n) \mid x_1 \in X_1 \text{ and } \cdots \text{ and } x_n \in X_n\}.$$

Fig. 5 The Cartesian product can be generalized to the **n-ary** Cartesian product over n sets $X_1, \ldots, X_n$

As for the anticipated tool for this isotope pattern calculator, the **Cartesian Algorithm** facilitates the output for this proposed software learning tool. From the current calculation technique used, however the other algorithms such as **Z-score Algorithm, Polynomial Algorithm and Binomial Algorithm** are scrutinized to observe the disparity each of the characteristics.

According to the source that refers to Cartesian product, Cartesian product is a direct product of sets and it is named after *René Descartes,* whose formulation of analytic geometry gave rise to this concept. Specifically, the Cartesian product of two sets X (for example the points on an x-axis) and Y (for example the points on a y-axis), denoted $X \times Y$, is the set of all possible ordered pairs whose first component is a member of X and whose second component is a member of Y. A Cartesian product of two finite sets can be represented by a table, with one set as the rows and the other as the columns, and forming the ordered pairs, the cells of the table, by choosing the element of the set from the row and the column [13].

Referring to *Rosen*, algorithm is a finite set instruction for performing a computation or solving a problem [14]. To perform the algorithm for the theoretical isotope generator, the concept of Cartesian product is applied. Cartesian product of sets $A_1, A_2 \ldots A_n$, denoted by $A_1 x A_2 \ldots A_n$ is the set of ordered n-tuples $(a_1, a_2 \ldots a_n)$, where a_j belongs to A_i for $i = 1, 2, \ldots, n$. In other words $A_1 x A_2 X \ldots A_n = \{a1, a2, \ldots, a_n\} \mid a_i \ni A_i$ for $I = 1, 2 \ldots n\}$[xx].

1) Calculation Step 1 (*Cartesian Product*)

As atom C is treated as a set. The elements of the sets (^{12}C & ^{13}C) represent the mass number which comprises elements isotopic composition. [**C= {C$_{12}$, C$_{13}$} where C$_{12}$ = C1 and C$_{13}$ = C2**]. On the other hand, atom H is treated as a set. The elements of the

sets (^{1}H & ^{2}H) represent the mass number which comprises elements isotopic composition. [**H= {H$_1$, H$_2$} where H$_1$ = H1 and H$_2$ = H2**] So, with the Cartesian product algorithm is used the isotopic distribution and exact value for the relative intensity can be obtained from the calculations for this proposed system. With the implementation of this algorithm, for example let say the combination of CH$_2$ (*methylene*) will deliver the results of eight different isotopes [C1 H1 H1; C1 H2 H1; C1 H1 H2; C1 H2 H2; C2 H1 H1; C2 H2 H1; C2 H1 H2; C2 H2 H2]:
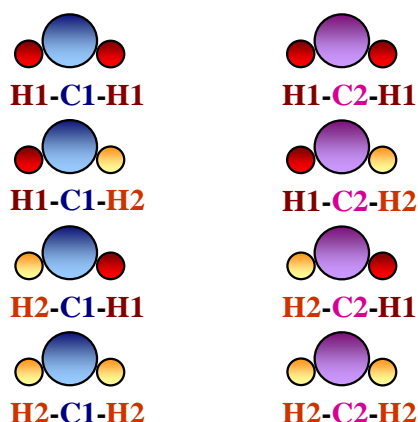

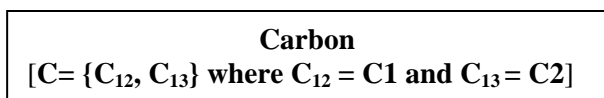Fig. 6 Calculation Step 1 (Cartesian Product): Possible combination of CH$_2$ isotopes

Carbon
[C= {C$_{12}$, C$_{13}$} where C$_{12}$ = C1 and C$_{13}$ = C2]

Fig. 7 Calculation Step 1 (Cartesian Product): Representation of the Carbon mass numbers

Hydrogen
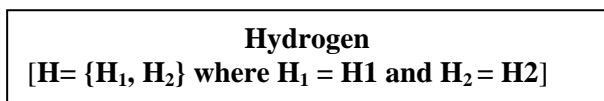[H= {H$_1$, H$_2$} where H$_1$ = H1 and H$_2$ = H2]

Fig. 8 Calculation Step 1 (Cartesian Product): Representation of the Hydrogen mass numbers

2) Calculation Step 2 (*Cartesian Product*)

Through the procedure Cartesian product (A1, A2,.....,An are atom in the molecule) the combination of atoms can be performed with [**j=1**

(**j<=n and j≠0) while combination={a$_1$,a$_2$,.....,a$_n$| a$_j$ Э A$_j$ j=1,2,3,...,n}**]. As proceed to other steps, the mass for each combination is calculated with [**Mass=a$_1$+a$_2$+.....a$_n$**] as the mass number for C is 12 and 13 whereas the mass number for H is 1 and 2.

[j=1 (j<=n and j≠0) while combination={a$_1$,a$_2$,.....,a$_n$| a$_j$ Э A$_j$ j=1,2,3,...,n}]

Fig. 9 Calculation Step 2 (Cartesian Product): Formula for the possible combination isotopes

[Mass=a$_1$+a$_2$+.....a$_n$]

Fig. 10 Calculation Step 2 (Cartesian Product): Calculation of the mass number of the combinations

3) Calculation Step 3 (*Cartesian Product*)

After we have detected the combination, the value of the mass number for each combination is added to find the exact value of mass number for each isotope [C1 H1 H1; C1 H2 H1; C1 H1 H2; C1 H2 H2; C2 H1 H1; C2 H2 H1; C2 H1 H2; C2 H2 H2] = [12 + 1+1 = 14; 12+1+2=15; 12+2+1=15; 12+2+2=16; 13+1+1=15; 13+1+2=16; 13+2+1=16; 13+2+2=17].
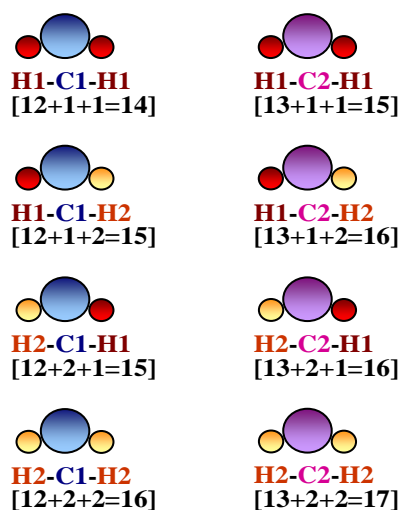

Fig. 11 Calculation Step 3 (Cartesian Product): Possible