On Solved and Unsolved Problems in Chemistry

Milan RANDIĆ*

National Institute of Chemistry, Ljubljana, Slovenia
*Visitor, Honorary Member of the Institute

(Received: January 30, 2017; Accepted for publication: Mach 1, 2017; Online publication: June 6, 2017)

In this article we have illustrated one solved problem in chemistry, which left room for improvements. This instance is raising an issue when a solved problem is to be considered solved. As we will see by using novel tools, not known at the time of solving a problem, one can arrived at additional unknown information on the problem. We will consider several problems of chemistry, some even considered as having no exact solution, which have been solved by using previously unknown concepts in chemistry. The last problem which we consider has been solved, but the problem is still open, as it might have additional solutions.

Keywords: Conjugated Circuits, Aromaticity, “QSAR nightmare”, Young diagrams, Graphical Bioinformatics

1 Introduction

The current article is dealing with a topics on "Mathematical chemistry. Look back and beyond," but I would like to stress here at the beginning that I did look back and beyond, but my interests have been about things that may be lying beyond rather than lying back. For some time I thought that it would be very interesting and thought-provoking to have, from time to time, a list of unsolved problems in mathematics, physics, chemistry, biology and medicine. But as far as I know, with a single exception, such lists were not available in the past nor are currently available. The exception happened at the turn of the last century, in 1900, when David Hilbert, one of the leading mathematicians of that time, was invited to deliver inaugural lecture at the the Second World Mathematical Congress held in Paris. He opened the Congress with a list of 23 unsolved problems of mathematics, which since that time have been in the center of attention of mathematicians for the next hundred years. The solution for one of the problems was reported already the next year, but it took more than hundred years till the last of the 23 problems has been solved. There is no doubt that finding solutions to these 23 unsolved problems of Hilbert have advanced mathematics visibly during the past 100 years. For almost forty years Hilbert himself was involved in working on the Hilbert Problem 6, which is:

Mathematical Treatment of the Axioms of Physics.

The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which already today mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.

As one can see this problem is about physics rather than mathematics.

In a recent book: Solved and Unsolved Problems in Structural Chemistry [1] the authors made the first step in the direction of searching for a list of 23 significant Problems in Chemistry, even if it may take 23 chemists, to arrive at it. They have proposed a Problem # 1 of chemistry, which is the last problem in the above book. Here we reproduce it:

Mathematical Treatment of the Axioms of Chemistry:

To treat … by means of axioms, those chemical sciences in which already today mathematics plays an important part; in the first rank are quantum chemistry, statistical mechanics, chemical graph theory, and graphical bioinformatics.

Hence, we have Problem Number 1, of our List of Unsolved Problems of Chemistry, which clearly has been inspired by the Hilbert Problem #6. Later we will add few problems hoping to be followed by other chemists and hopefully before not too
long jointly to arrive at a list of unsolved 23 important problems of chemistry.

2 A Solved Problem Revisited

We start by considering a solved problem, in order to illustrate that sometimes "solved" may be a relative, and not an absolute concept. Consider the set of Newton's equations:

\begin{align*}
    s_1 + c_1 &= 0 \\
    s_2 + c_1s_1 + 2c_2 &= 0 \\
    s_3 + c_1s_2 + c_2s_1 + 3c_3 &= 0 \\
    s_4 + c_1s_3 + c_2s_2 + c_3s_1 + 4c_4 &= 0 \\
    &\ldots
\end{align*}

Here we use the present day notation of matrices, which were unknown in the time of Newton, and will illustrate it on graphs, which also have not been known in the time of Newton. So \( s_k \) are the traces of \( A_k \), where \( A_k \) is the adjacency matrix \( A \) of a graph on \( k \)-th power, and \( c_k \) are the coefficients of the characteristic polynomial of a graph. Observe the structural simplicity of Newton's equations. They have simple pattern and allow one to write down the equation for \( s_k \) for any \( k \), without need to know the equations for lower values of \( k \).

The unsolved problem was to write the equations for the coefficients \( c_k \) of the characteristic polynomial in terms of the \( s_k \), the traces of \( A_k \). This problem remained unsolved for about 250 years and was finally solved in 1986, by Barakat [2]. Barakat has written computer program that gives the coefficients of the characteristic polynomial in terms of traces of powers of adjacency matrices. After several initial simpler equations soon the coefficients in the equations became very involved, as is illustrated below by the expressions for the coefficients \( c_9 \) and \( c_{10} \):

\begin{align*}
    c_9 &= (40,320 s_9 - 25,920 s_7s_2 - 20,160 s_6s_3 - 18,144 s_5s_4 + 9,072 s_5s_2^2 + 15,120 s_4s_3s_2 + 2,240 s_3^3 - 2,520 s_3s_2^3)9! \\
    c_{10} &= (-362,880 s_{10} + 226,800 s_8s_2 + 172,800 s_7s_3 - 15,100 s_6s_4 - 75,600 s_5s_2^2 + 72,576 s_4^2)10! \\
    &+ 120,960 s_8s_3s_2 - 56,700 s_7^2s_2 - 50,400 s_6s_3^2 + 18,900 s_5s_2^3 \\
    &+ 25,200 s_4^3s_2^2 - 954 s_3^5/10!\end{align*}

Hence, Barakat has solved the problem, and if needed his program could generate solutions for higher coefficients. However, as one can see, in contrast to the Newton's equations, here it appears impossible to find regularities in the resulting coefficients.

I have learned about the work of Barakat, which was published in Theoretica Chimica Acta, as the Editor of that journal asked me to review the article. I have recommended publication and mentioned to the author that the above cumbersome coefficients have some mathematical structure and can be expressed relatively simply in terms of Young diagrams. On the basis of my finding I can ensure everybody that his computer program gives correct answers and that there are no even typing errors in the reproduced results - though I neither checked the program nor I have an alternative program to check the results. Moreover, by finding the mathematical structure of the coefficients of \( A_k \) not only that I could check for accuracy, but if needed I could write down the expressions for higher coefficients if necessary, without use of computer and without use of recursion expressions, which I believe computer program has used.

I got a letter from the Editor of Theor. Chim. Acta, Professor Ruedenberg, who told me that I don't need to tell the author my findings, but that I should wait till his paper is published and then send my results for publication. This I did and my results based on Young diagram have appeared next year in J. Math. Chem [3]. The question is: How I was able to find mathematical regularity in the resulting coefficients? It has not been accidental. Consider, for example the second term in the expression for \( c_9 \), which is 25,920 \( s_7s_2 \). The question that I considered was: what has 25,920 to do with numbers 9, and its partition 7 and 2, which are the subscripts related to 25,920. It turns out that \( 9!/(7 \times 2) = 25,920 \). Partitions of numbers can be represented graphically by Young diagrams. For more see refs [1,3]. The point here that is important is that to find the above solution one has to have heard of Young diagrams, which are topics of higher algebra and discrete mathematics. If one does not know the "language" of Young diagrams one has no chance to find the above solution.
3 Solutions Found by Finding New "Languages" First

Let us continue by describing several solved problems, all of which were waiting for quite long time to be solved. Some may have been even considered as not having exact solution – but as we will see, they do. Because in each case the solutions were arrived at by use of novel concepts not known or not considered in chemistry prior to that time, the available solutions for most chemists who have not heard of the novelties involved, remained unknown. Hence, these problems qualify to be on the non-existing list of 23 unknown problems of importance to chemistry. Let us pretend that they are not solved, because too few chemists may know about them, and we will give them numbers Problem 2 to Problem 6. Here they are:

4 Problem Number 2

The Hückel $4n+2$ Rule of Aromaticity, which tells that monocyclic conjugated ring systems are aromatic if rings have $4n+2$ π-electrons, is too well known [4]. It well deserves to be so, because Hückel succeeded to explain why benzene $C_6H_6$ is aromatic and cyclo-octatetraene $C_8H_8$ is not – which was a great mystery up to that time. Still an open problem was to find generalized Hückel Rule of Aromaticity valid for polycyclic conjugated systems. The problem has been solved in 1977, but how many chemists know this? The key novelty for successful solution of this problem, reported in [5], have been conjugated circuits, found in 1975 in Kekulé valence structures, 110 years after August Kekulé proposed the ring structure for benzene. The solution is simple: The polycyclic conjugated systems having only $4n+2$ conjugated circuits are aromatic.

5 Problem Number 3

It is well known that molecular descriptors used in multiple regression analysis (MRA) all mutually inter-correlated, some more some less. Regardless whether one uses mathematical descriptors or selected physicochemical properties as descriptors, what happens is that when adding a new descriptor in MRA in order to reduce the standard error of MRA, the coefficients of previous regression equations change, some more or less. This depends on the degree of inter-correlation of descriptors used. This has been known as "QSAR nightmare." The problem of construction of orthogonal set of vectors was solved over 100 years ago, but generalization of such problem for MRA was not known. However in 1991 this problem was solved. As outlined in [1] and [6] by using residues of correlations between the descriptors and correlation between the residuals one can construct set of orthogonal descriptors that result in stable sets of MRA equations.

6 Problem Number 4

One of the central problems of Bioinformatics, which is a hybrid discipline involving chemistry, physics, mathematics, computer science and informatics, has been the problem of protein and DNA alignments. The first computer program solving this problem emerged in 1970 [7]. Currently the most widely used program, BLAST [8,9], is about 25 years old. All computer programs give approximate solutions. The problem is: Is there an exact solution to the protein alignment problem? Exact solutions are solutions obtained without using empirical parameters and trial-and-error shifting of the sequences considered. Very recently, in 2014, (see [1] and [10]) an exact solution to the protein alignment problem was reported. The novel "languages" used here were the amino acid adjacency matrices and their modification. In 2016 was published a comparison of BLAST and the exact solution on a pair of proteins having 170 amino acids, for which BLAST found 89 pairing while the exact solution gives 95 [11].

7 Problem Number 5

Is it possible to formulate mathematically rigorous definition of aromaticity? This is not an easy problem. Generations of chemists have found the concept of aromaticity useful despite that it has never been clearly defined. We will here restrict the problem of aromaticity to the class of benzenoid and non-benzenoid hydrocarbons, and will not consider heterocyclic compounds that have additional factors that may play an important role for aromaticity. The generalized Hückel $4n+2$ Rule (Problem # 2) already told us which polycyclic conjugated hydrocarbons are aromatic and which are at best partially aromatic. However, here we want a quantitative definition of aromaticity, which will for any pair of compounds tell, which molecules
is more aromatic and which is less. The sought definition of aromaticity should not use empirical information, but solely be based on structural information. So far such a solution has not been reported satisfying stated conditions, but that does not mean that there is none.

8 Problem Number 6

This problem relates to the Hückel 4n+2 Rule of Aromaticity, which says that monocyclic systems having \(4n+2\) π-electrons are aromatic. The unsolved problem is: Are there additional classes of molecules for which one can find that they are aromatic? One solution to this problem has been found in 1989, but this problem may have additional solutions, yet to be found.

In 1989 Gutman and Cyvin [12] have shown that all benzoid hydrocarbons, which are systems built solely from fused benzene rings, are aromatic, regardless how many benzene ring and how many π-electrons they may have. Observe how more important in practice is this result, which applies to a class of hundreds and hundreds of molecules, in comparison to the original Hückel Rule, which applies in practice to about a dozen of molecules. However, this does not make the Hückel 4n+2 Rule less important – on the contrary because this result relates to Generalized Hückel 4n+2 Rule, it makes the Hückel 4n+2 Rule even more important.

9 Concluding Remarks

We hope that this article may stimulate search for additional unsolved problems in chemistry, and even solved problems, if they have not been openly accepted or rejected by current experts in the corresponding areas. By including problems 2–6 in the list of Unsolved Problems of Chemistry is justified so far because NO PUBLIC COMMENTS yet have been made by anyone that the solutions are accepted as correct. So far there have been only claims made by authors that the problems have been solved! That is not enough. Authors of the listed problems may have private comments, as is the letter shown below from the year 1976, about Problem #2 on solution of the Generalized Hückel 4n+2 Rule, but this is from private correspondence and not a public source.

Soon after the paper on aromaticity and conjugation [5] was published Professor Helwinkel, Department of Chemistry, the University of Heidelberg, Germany sent this letter to the author:

"I just came from reading your most interesting paper on Aromaticity and Conjugation (JACS 99 (1977) 444). I think that your extension of the Hückel Rule provides a most important improvement in the understanding of the properties of polynuclear conjugated systems. This even more so because your method of dividing fused systems formally into cyclic conjugated sub-entities can be understood and performed by every student without sophisticated theoretical/mathematical background. Personally I think that always these theories are the best ones which allow simple quantifications of inherent chemical intuitions. And this is exactly the case with your elegant substructure counting and weighting procedure. I am sure that these ideas will find very broad (and grateful) acceptance in chemical world . . ."

Despite the optimism of Professor Helwinkel, let me remind the readers of Aristotle’s saying: One swallow does not make a Spring. We need more swallows in science, but they do not appear in sight yet, particularly not covering the area of mathematical chemistry, which remains outside attention of most chemists at large.

This article is dedicated to Professor Roald Hoffmann, first and the foremost theoretical chemist, also writing poetry, plays, and nonfiction on art and science, on the occasion of his 80th anniversary.

References

[1] M. Randić, M. Nović, D. Plavšić, Solved and Unsolved Problems in Structural Chemistry, CRC Press, Boca Raton, FL (2016).
[2] R. Barakat, Theor. Chim. Acta, 69, 35 (1986). DOI:10.1007/BF00526290
[3] M. Randić, J. Math. Chem., 1, 145 (1987).
[4] E. Hückel, Grundlehre der Theorie ungesättiger und aromatischer Verbindungen. Berlin: Verlag Chemie, pp. 77–85 (1938).
[5] M. Randić, J. Am. Chem. Soc., 99, 444 (1977). DOI:10.1021/ja00444a022
[6] M. Randić, Chem. Inf. Comput. Sci, 31, 311 (1991). DOI:10.1021/ci00002a018
[7] S. B. Needleman, C. D. Wunsch, *J. Mol. Biol.*, **48**, 443 (1970). PMID:5420325, DOI:10.1016/0022-2836(70)90057-4

[8] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.*, **215**, 403 (1990). PMID:2231712, DOI:10.1016/S0022-2836(05)80360-2

[9] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.*, **25**, 3389 (1997). PMID:9254694, DOI:10.1093/nar/25.17.3389

[10] M. Randić, *J. Comput. Chem.*, **33**, 702 (2012). PMID:22213065, DOI:10.1002/jcc.22903

[11] M. Randić, T. Pisanski, *J. Comput. Chem.*, **36**, 1069 (2015). PMID:25800773, DOI:10.1002/jcc.23892

[12] I. Gutman, S. J. Cyvin, *J. Mol. Struct. THEOCHEM*, **184**, 159 (1989). DOI:10.1016/0166-1280(89)85141-3