Development and automation of the algorithm for determining the enthalpy of formation for organic compounds

A I Akhmetyanova

Department Mathematics and Information Technologies, Bashkir State University, 32, Zaki-Validi, Ufa, 450076, Russia

E-mail: ai-albina@mail.ru

Abstract. The best solution for the development of chemical and technical processes is possible if there is reliable information on the physicochemical and thermodynamic properties of chemical compounds only. An urgent task is the development of universal theoretical methods for calculating and predicting the properties of chemical compounds, which make it possible to independently assess the energy content of a substance, identify incorrect data and, ultimately, reliably determine the thermal effects of chemical reactions, which is important for the development of technological regulations and conscious control of chemical processes. The work is devoted to the description of the mathematical apparatus for constructing the basis of homodesmic reactions on the example of cyclic compounds.

1. Introduction

Thermodynamic calculations are a common and important part of many chemical research. They form the basis for the development of highly efficient industrial technologies for the synthesis of compounds. Consequently, there is need to collect reliable information on the thermodynamic properties of individual substances. Experimental data on thermodynamic properties are available for a limited number of compounds; therefore, the development and improvement of methods for calculating thermodynamic quantities is an important and urgent problem. The development of informatics and the development of new methods of quantum chemistry make it possible to carry out theoretical calculations of the thermodynamic properties of chemical compounds with an accuracy comparable to the errors of experimental studies. This allows the development of additive methods that can be based not only on experimental data, but also on reliable theoretical data. Therefore, it becomes possible to establish regularities linking the thermodynamic properties of substances with their structure for a variety of compounds. An approach based on graph theory [12] is proposed, which became the basis for creating an algorithm for calculating the ratio of parameters when constructing participants in homodesmic reactions.

2. Materials and methods

An optimal solution of the problems that may arise in the development of a chemical-technological process is possible only if there is reliable information on the chemical characteristics and thermodynamic
properties of chemical compounds. A significant amount of this data has been accumulated. However, the rapid development of technology and the emergence of an increasing number of new chemical compounds create a gap between the need for data and their availability. Experimental determination of the thermodynamic values of organic compounds is often difficult and sometimes impossible. In addition, among this huge amount of experimental information, there are also unreliable data. Therefore, the urgent task of creating universal and high-precision theoretical methods for calculating and predicting the properties of chemical compounds has emerged, which makes it possible to independently estimate the energy content of a substance, identify erroneous data and, finally, reliably determine the thermal effects of chemical reactions, which is of decisive importance for the development of technologies and quantitative control of chemical activity.

In practice, reaction mechanisms often involve several elementary steps and the substances involved. In the chemical analysis of the method, it is necessary to solve a system of differential equations, where each equation determines the conversion rate of one of the respondents. In general, solving such a problem is a very laborious process. Moreover, it is impossible to measure the concentrations of all substances involved in the mechanism.

Chemical thermodynamics allows you to calculate the heat effect of a reaction, determine whether a particular reaction is possible, and calculate it's equilibrium state, that is, the limit to which it can proceed. To calculate the thermal effect of a reaction, it is necessary to know the enthalpies of formation of all the substances involved. Often, experimental data on the enthalpy of formation taken from different sources have different meanings. Moreover, obtaining experimental data for some compounds is difficult, prohibitively expensive, or even impossible. These problems prompted the emergence of theoretical methods for assessing the enthalpy of formation of chemical compounds.

One of these methods is based on the use of homodesmic reactions (HDR) [9]. The HDR method is the decomposition of the initial compound into internal thermochemical groups [1]. It is clear that the degree of such decomposition depends on the number of internal groups that make up the chemical compound. The totality of all independent HDRs constitutes the core of the HDR. Therefore, having obtained different estimates of the enthalpy of formation for each HDR, one can judge the reliability of experimental data from different sources. Isodesmic and homodesmic concepts have been successfully applied to describe the thermochemical properties of both individual molecules and some classes of organic compounds [17-20].

3. Formulation of the problem

We will test the presented HDR method using the example of two cyclic compounds cis-1,3-dimethylcyclobutane \( \text{C}_6\text{H}_{12} \) and trans-1,3-dimethylcyclobutane \( \text{C}_6\text{H}_{12} \) (figure 1). Cis-trans isomerism is one of the types of stereoisomerism: it consists in the possibility of the arrangement of substituents on one or opposite sides of the plane of a double bond or a non-aromatic ring.

![Molecules cis-1,3-dimethylcyclobutane C_6H_{12} and trans-1,3-dimethylcyclobutane C_6H_{12}.](image-url)

**Figure 1.** Molecules cis-1,3-dimethylcyclobutane \( \text{C}_6\text{H}_{12} \) and trans-1,3-dimethylcyclobutane \( \text{C}_6\text{H}_{12} \).
Applying Benson's method [1] with the condition of continuity of the multiple bond C = C, we represent the structure of 1,3-dimethylcyclobutane as a combination of two types of internal (B01, B02) and terminal (K01) groups (figure 2).

![Figure 2](image)

Figure 2. C₆H₁₂ 1,3-dimethylcyclobutane molecule presented as a combination inner and end groups.

Next, we will present all possible variants of the "unfolded" molecule, which is obtained from the initial one by "breaking" the cyclic bond (figure 3).

![Figure 3](image)

Figure 3. All possible variants of cyclic bond breaks in the 1,3-dimethylcyclobutane C₆H₁₂ molecule.

Figure 3 shows that one break is enough to break the cycle in the molecule. In addition, the breaks «1-2-3-4» are symmetrical, so in our calculations we will use only one of the four breaks. By analogy, from the options «5-6», «7-8», «9-10», as well as «11-14» gaps presented in the figure, we will take into account only one option. Thus, we have obtained 6 basic HDRs for the 1,3-dimethylcyclobutane molecule.

1. C₆H₁₂ + C₂H₆ (Ethane) → C₈H₁₈ (2,4- Dimethylhexane) «1-2-3-4»
2. C₆H₁₂ + 2 C₂H₆ (Ethane) → 2 C₅H₁₂ (2- Methylbutane) «5-6»
3. C₆H₁₂ + 2 C₂H₆ (Ethane) → C₄H₁₀ (Isobutane) + C₆H₁₄ (3- Methylpentane) «7-8»
4. C₆H₁₂ + 2 C₂H₆ (Ethane) → C₇H₁₆ (2,4- Dimethylpentane) + C₃H₈ (Propane) «9-10»
5. \( \text{C}_6\text{H}_{12} + 3 \text{C}_2\text{H}_6 \rightarrow \text{C}_4\text{H}_{10} + \text{C}_5\text{H}_{12} \) (Ethane) \( \rightarrow \text{C}_4\text{H}_{10} \) (Isobutane) \( + \text{C}_3\text{H}_8 \) (Propane) \( \text{«11-14»} \)

6. \( \text{C}_6\text{H}_{12} + 4 \text{C}_2\text{H}_6 \rightarrow 2 \text{C}_4\text{H}_{10} + 2 \text{C}_3\text{H}_8 \) (Ethane) \( \rightarrow \text{C}_4\text{H}_{10} \) (Isobutane) \( + 2 \text{C}_3\text{H}_8 \) (Propane) \( \text{«15»} \)

4. **Software implementation**

The process of constructing the HDR base is automated. The development of the program is based on the graphic interpretation of chemical compounds [8, 10]. The use of graph theory makes it possible to automate the procedure for identifying the base of the HDR, as well as to obtain a visual geometric interpretation of the base, which is important for subsequent physico-chemical analysis.

The relational database was developed to store information on the energy characteristics of chemical compounds [7]. The paper [2] analyzes the possibilities of using DBMS of various classes for storing and searching data on the thermodynamic properties of individual substances. The results of the analysis show that the use of relational DBMSs for storing a large amount of information about the physical and chemical properties of substances can be considered quite acceptable. For acyclic compounds, the program was developed earlier by the authors in [3, 4, 5].

The use of adjacency matrices is one of the methods of representing a chemical graph as a matrix. In the studied example of a cyclic 1,3-dimethylcyclobutane molecule, the adjacency matrix will have the following form (figure 4):

![Figure 4](image)

**Figure 4.** \( \text{C}_6\text{H}_{12} \) 1,3-dimethylcyclobutane molecule and its adjacency matrix.

In the adjacency matrix, ones and zeros denote the relationships between the inner and end groups. For clarity, the program will illustrate the first option for breaking one bond, from Figure 3, between groups B01 and B02 (figure 5).

![Figure 5](image)

**Figure 5.** Breaking the connection between the groups B01 and B2 and the extended adjacency matrix.
When breaking any bond in the connection, we must attach a suitable end group to each inner group, as shown in Figure 5. Therefore, the adjacency matrix must increase by two columns and two rows, since two K01 leaf groups are added. The program performs similar actions for the rest of the gaps. Then initialization from the database takes place and the calculations take place.

The implemented software [7, 8] constructs the HDR basis, calculates the formation enthalpies of the chemical compound under study (figures 6-7). All available experimental data were obtained from NIST Standard Reference Database Number 69 [6]. A specialized enumeration method was applied, which was described by us earlier [11], based on recursion.

Also, a database was created and registered to store information on chemical compounds [7]. To design a database, the MySQL database management system is used, which was described by us in the following articles [10, 15, 16].

![Figure 6](image.png)

**Figure 6.** The program demonstrating the calculation for the compound of trans-1,3-dimethylcyclobutane.

The program that implements the algorithm for identifying the GDR basis and determining the enthalpies of formation of chemical compounds was developed in the object-oriented programming
language Delphi. The homodesmic method, implemented in the program code, makes it possible to determine such energy characteristics as enthalpies of formation, cycle stress energies, non-bonded interactions energy, etc. The program has been tested on a representative set of chemical compounds, there is a certificate of registration of the computer program[8].

The selection of the basic set of GDR in the form of group separation reactions allows performing transformations over the adjacency matrix of the vertices of the molecular graph of a chemical compound. It unambiguously describes a chemical compound, and thus allows the algorithm to be translated into machine language. The molecular graph of a chemical compound is transformed into a graph, the vertices of which are groups, the edges are group bonds. The graph uniquely describes the vertex adjacency matrix.

A mathematical apparatus has been developed for assessing the energy characteristics of chemical Therefore, this work presents a graph-theoretical interpretation of the structure of cis-1,3-dimethylcyclobutane and trans-1,3-dimethylcyclobutane molecules. On the basis of the algorithm, the
basis of homodesmic reactions for the compounds under study was constructed. At the moment, we are faced with the task of expanding the database and testing the resulting software for various organic compounds in order to identify shortcomings and eliminate them in the software.

5. Results and discussion
The results of the implementation of the mathematical apparatus for assessing the energy characteristics of chemical compounds were: analysis of experimental data on the enthalpies of formation of organic radicals for a representative set of compounds for the purpose of further analysis; quantum-chemical calculations for a test set of compounds using different levels of theory, to form an analyzed array of absolute enthalpies of radical particles; graph-theoretic analysis of the compounds of the test set, which makes it possible to form the HDR basis for each free organic radical; theoretical calculation of the energy content of the compounds of the test set, identification of poorly substantiated experimental data, refinement of the values of the enthalpies of formation of free radicals; calculation of the dissociation energy of chemical bonds using the found values of the enthalpies of formation for free radicals; completion of software for theoretical methods of calculating and predicting the properties of chemical compounds.

The program has been tested on a representative set of various cyclic organic compounds. The stress energies of the cycle of cyclic structures are calculated using the homodesmic approach, the influence of substituents and non-bonded effects on the value of the cycle stress energy is considered, for bicyclic compounds the influence of the structure rigidity on the value of the cycle stress energy is considered. Experimental data on the enthalpies of formation of chemical compounds are analyzed. A comparison is made with the experimental data obtained for a number of structures under study, which showed a high degree of convergence of the calculated and experimental data. The values of the thermal effects of the HDR are calculated. It was found that the thermal effect is the sum of the cycle stress energy and the energies of unaccounted for non-bonded interactions (gauche effect, intramolecular hydrogen bond); the thermal effect of HDR is also influenced by the effect of small molecules, the nature of which is presumably associated with a small dimensional inconsistency of the composite calculation method. Separate determination of all the listed effects is carried out. The obtained values of the energies of stress cycles are correlated with calculations using experimental data on the enthalpies of formation of compounds participating in the HDR. It has been established that the proposed calculation method is characterized by a high accuracy in determining the thermochemical parameters of cyclic organic compounds [13, 14]. Based on the obtained cycle stress energies, the dependence of this value on the structural features of the compound is analyzed. It is concluded that the values of the stress energy of the cycle are not influenced by alkyl, alkenyl substituents, and it is also shown that for annelated and spiro-joined polycycles, the stress energy of the cycle is non-additive. Therefore, the homodesmic method for determining non-valence effects for organic compounds can be considered as an alternative to expensive and time-consuming experimental methods.

The program displays all possible ways of decomposition of the analyzed compound, the subsequent choice of the necessary reactions by the user depends on the set physical and chemical tasks. In particular, the homodesmic approach is applied to calculate the enthalpy characteristics of radical structures. A critical analysis of the literature in the field of theoretical methods for determining the enthalpies of formation of hydrocarbon radicals has been carried out, an array of experimental and calculated values of the enthalpies of formation has been accumulated. When choosing the best type of HDR, the effect of long-range interaction of an unpaired electron with neighboring atoms was taken into account, the minimum size of a thermochemical radical group was determined, which allows calculating the enthalpies of formation of radicals with the accuracy of the best experimental techniques. For the test set of alkyl
radicals, the basic sets of HDR were constructed, and the thermal effects of HDR were calculated using the quantum-chemical method G4.

It was found that the construction of a thermoneutral “ideal” HDR and high-precision determination of the enthalpies of formation of radicals is possible when radical compounds are used as reference structures, i.e. the careful execution of the isogyric balance of the HDR is very important. In particular, the use of an ethyl radical as a coagent is impractical, while the use of an n-propyl radical significantly improves the results of calculating the enthalpies of formation of primary alkyl radicals, the error in determining the enthalpies of formation is no more than 1 kJ / mol.

6. Conclusions
A mathematical apparatus has been developed for assessing the energy characteristics of chemical compounds (enthalpy of formation, cycle stress energy, bond strength) [11]. A comparative calculation method based on the use of homodesmic reactions (HDR) was applied. In the general case, for an arbitrary chemical compound, there is an ambiguity in the choice of the HDR. This problem was solved using the procedure of decomposition of the molecular graph of the compound under study and the corresponding matrix of bonds of thermochemical groups. The result of the decomposition is the basis of independent HDRs. Determination of the HDR basis allows one to independently assess the energy of molecules, control the reproducibility of results, and increase the reliability of the theoretical determination of energy characteristics.

Acknowledgements
This work was supported by the Russian Foundation for Basic Research, and the Government of the Republic of Bashkortostan as part of scientific project No. 18-07-00584

References
[1] Benson S.W. 1971 Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters (Wiley, New York) p 308
[2] Belov G.V. 2003 Using a DBMS for storing and retrieving information on the physical and chemical properties of substances 1 p 3 (in Russian)
[3] Khursan S L Ismagilova A S Spivak S I 2017 A graph theory method for determining the basis of homodesmic reactions for acyclic chemical compounds Doklady Physical Chemistry 2(474) pp 99-102
[4] Akhmerov A A Ismagilova A S Spivak S I Khursan S L 2015 Construction of homodesmic reactions and calculation of the enthalpy of formation of organic compounds Certificate of state registration of the computer program No. 2015617060 (in Russian)
[5] Akhmerov A A Ismagilova A S Spivak S I Khursan S L 2015 Energy characteristics of organic compounds Certificate of state database registration No. 2015621003(in Russian)
[6] Afeefy H Y Liebman J F Stein S E Neutral Thermochemical Data In: NIST Chemistry Webbook, NIST Standard Reference Database Number 69. Eds. P.J. Linstrom and W.G. Mallard, National Institute of Standards and Technology, Gaithersburg Md, 20899, DOI: https://doi.org/10.18434/T4D303
[7] Akhmetyanova A I Ziganshina F T Ismagilova A S Akhmerov A A 2020 Energy characteristics of organic compounds for calculating the standard enthalpy of formation and stress energy of cycles Certificate of state. database registration No. 2020621607(in Russian)
[8] Akhmerov A A Ziganshina F T Akhmetyanova A I Ismagilova A S Yunusov A A Theoretical calculation of energy characteristics of organic compounds Certificate of state. registration of a computer program No. 2020660354 (in Russian)
[9] Khursan S L Ismagilova A S Akhmerov A A Spivak S I 2016 Constructing homodesmic reactions for calculating the enthalpies of formation of organic compounds Russian Journal of Physical Chemistry A 4 (90) pp 796-802

[10] Akhmetyanova A I Ismagilova A S Ziganshina F T 2019 Mathematical apparatus for the construction of homodesmic reactions Proceedings - 2019 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency pp 69-72

[11] Khursan S L Ismagilova A S Akhmetyanova A I 2018 Determining the basis of homodesmotic reactions of cyclic organic compounds by means of graph theory Russian Journal of Physical Chemistry A 7 (92) pp 1312-1320

[12] Ismagilova A S Khamidullina Z A Spivak S I Morozkin N D 2019 Graph-theoretic interpretation of inverse problems of chemical kinetics High Energy Chemistry 6 (53) pp 421-424

[13] Spivak S I Ismagilova A S Akhmerov A A 2015 Computer analysis of the graphs of complex chemical reactions High Energy Chemistry 4 (49) pp 217-222

[14] Pedley J B Naylor R D Kirby S P 1986 Thermochemical Data of Organic Compounds (Chapman & Hall, London) p 792

[15] Ziganshina F T Akhmetyanova A I Ismagilova A S Khursan S L Akhmerov A A 2018 Graph-theoretic method for determining homodesmic reactions for cyclic chemical compounds Control systems and information technologies 4 (74) pp 72-76 (in Russian)

[16] Ziganshina F T Ismagilova A S Akhmetyanova A I Akhmetshina E S Akhmerov A A 2019 Computer modeling of the problem of determining the basis of homodesmic reactions Control systems and information technologies 4 (78) pp 10-15 (in Russian)

[17] Khursan S L 2004 The Standard Enthalpies of Formation of Fluorinated Alkanes: Nonempirical Quantum-Chemical Calculations Russian Journal of Physical Chemistry A 1 (78) pp S34-S42

[18] Dorofeeva O V Ryzhova O N Moiseeva N F 2008 The quantum-chemical determination of group contributions to the thermodynamic properties of organophosphorus compounds Russian Journal of Physical Chemistry A 6 (82) pp 933-937

[19] Dorofeeva O V Ryzhova O N Suntsova M A 2013 Accurate prediction of enthalpies of formation of organic azides by combining G4 theory calculations with an isodesmic reaction scheme Russian Journal of Physical Chemistry A 117 (32) pp 6835-6845

[20] Dorofeeva O V Osina E S 2017 Performance of DFT, MP2, and composite ab initio methods for the prediction of enthalpies of formations of CHON compounds using isodesmic reactions Computational and Theoretical Chemistry 1106 pp 28-35