

## **МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ**

УДК 004.021, 004.043, 004.4

### **РАЗРАБОТКА АЛГОРИТМА И МЕТОДА ТРАНСФОРМАЦИИ ЗАПИСИ АТОМНО-МОЛЕКУЛЯРНЫХ СИСТЕМ**

*Статья поступила в редакцию 02.05.2022, в окончательном варианте – 16.05.2022.*

**Смирнова Юлия Александровна**, Астраханский государственный университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,

старший преподаватель кафедры информационной безопасности и цифровых технологий, ORCID: 0000-0002-3807-5062, e-mail: yuliya\_smirnova\_2048@mail.ru

**Головацкая Леся Ивановна**, Каспийский институт морского и речного транспорта имени генерал-адмирала Ф.М. Апраксина – филиал Федерального государственного бюджетного образовательного учреждения высшего образования «Волжский государственный университет водного транспорта», 414000, Российская Федерация, г. Астрахань, ул. Никольская, 6,

кандидат технических наук, доцент, профессор кафедры математических и естественнонаучных дисциплин, ORCID: 0000-0002-7181-2845, e-mail: lesy\_g@mail.ru

В статье описывается алгоритм и метод трансформации записи атомно-молекулярных систем. Приведен перечень различных форм записей, проведен анализ структуры записей и актуализирован процесс разработки программы для конвертации Z-матриц в различные форматы и формы записи данных в файл. В процессе работы рассматриваются и используются форматы данных в квантово-химических программах: Gaussian, GAMESS, MOPAC. В результате исследования получено свидетельство о регистрации программы для ЭВМ. Программа может использоваться в учебном процессе на кафедрах химии, биологии, а также для лаборатории математического моделирования в химических областях. Функциональные возможности: программа осуществляет связь между квантово-химическими программами. Например, MOPAC помогает нарисовать молекулу и выгрузить данные её построения. На основе данных программа формирует Z-матрицу внутренних координат и записывает её в текстовый файл .txt, в котором уже задан код запуска расчета устойчивой структуры в GAMESS. Рассмотренный подход позволяет повысить эффективность в создании записей атомно-молекулярных систем, проведении сравнительных и научных квантово-химических вычислений, сокращает время на определение устойчивой структуры молекулярных компонентов и получение готовой Z-матрицы.

**Ключевые слова:** алгоритм, атомно-молекулярные системы, Gaussian, GAMESS, MOPAC, трансформация, конвертация

### **DEVELOPMENT OF AN ALGORITHM AND METHOD OF RECORD TRANSFORMATION OF ATOMIC AND MOLECULAR SYSTEMS**

*The article was submitted to the editorial office 02.05.2022, in the final version – 16.05.2022.*

**Smirnova Yulia A.**, Astrakhan State University, 20a Tatishchev St., Astrakhan, 414056, Russian Federation, Senior Lecturer, Department of Information Security and Digital Technologies, ORCID: 0000-0002-3807-5062, e-mail: yuliya\_smirnova\_2048@mail.ru

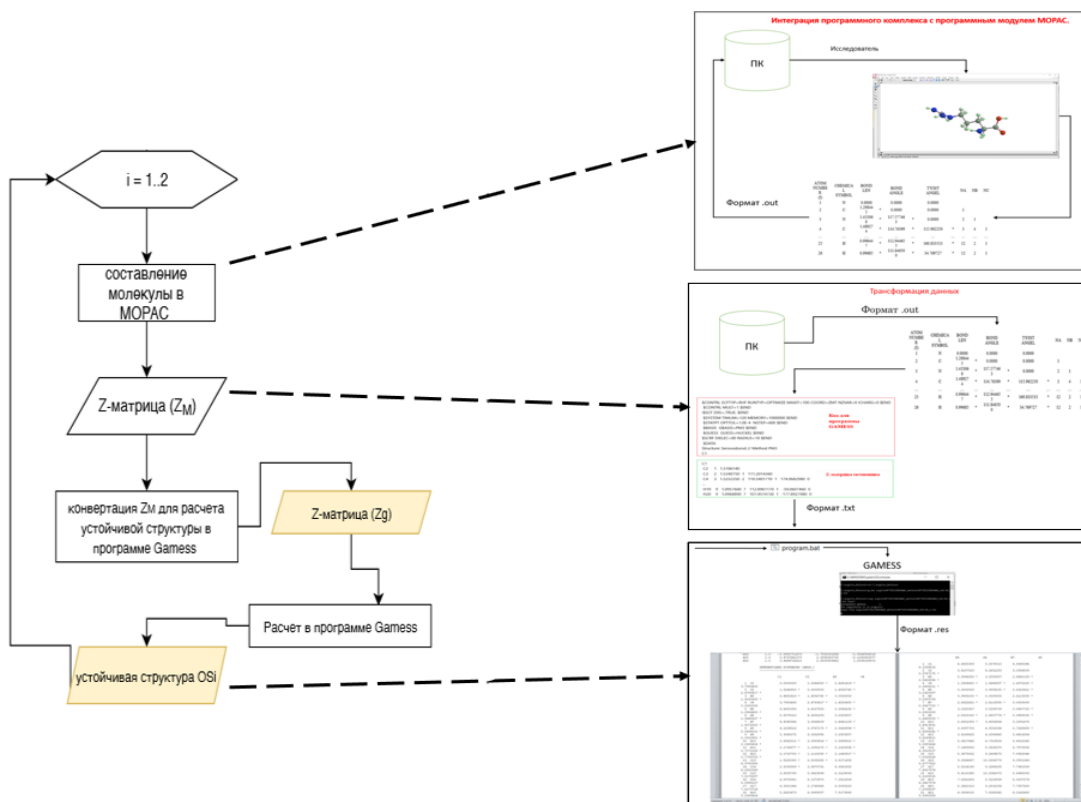
**Golovatskaya Lesya I.**, Caspian Institute of Sea and River Transport named after general-admiral F.M. Apraksin – branch of the Federal State Budgetary Educational Institution of Higher Education "Volga State University of Water Transport", 6 Nikolskaya St., Astrakhan, 414000, Russian Federation,

Cand. Sci. (Engineering), Associate Professor, Professor of Mathematical and Natural Sciences Department, ORCID: 0000-0002-7181-2845, e-mail: lesy\_g@mail.ru

The article describes algorithm and method of transformation of atomic-molecular systems records. The list of different record forms is given, the analysis of record structure is carried out and the process of program development for conversion of Z-matrices into different formats and forms of data records in a file is actualized. In the process, the data formats in quantum-chemical programs are reviewed and used: Gaussian, GAMESS, MOPAC. As a result of the research, a certificate of registration of a computer program was obtained. The program can be used in the educational process at the departments of chemistry, biology, as well as for the laboratory of mathematical modeling in chemical areas. Functionality: the program makes a link between quantum-chemical programs. For example, MOPAC helps to draw a molecule and to upload its construction data. On the basis of the data the program forms Z-matrix of internal coordinates and writes it in a text file .txt, in which the code of start of calculation of stable structure in GAMESS is already set. The considered approach allows to increase efficiency in creation of records of atomic-molecular systems, to carry out comparative and scientific quantum-chemical calculations, to reduce time for definition of stable structure of molecular components and to receive ready Z-matrix.

**Keywords:** algorithm, atomic-molecular systems, Gaussian, GAMESS, MOPAC, transformation, conversion

## Graphical annotation (Графическая аннотация)



**Введение.** Изучение веществ на атомном уровне в настоящее время широко развивается, что подтверждает ряд работ в данном направлении в различных областях науки [1–5]. Такое изучение не может проходить без молекулярного моделирования, часто называемого атомистическим моделированием. Опыт моделирования каких-либо процессов предполагает необходимость использования различных программных инструментов [6–7]. Каждый программный продукт использует свои шаблоны во входных и выходных документах. Так, например, в биохимии и биоинформатике существуют различные способы описания структуры молекулы. Каждое описание может отличаться в различных программах.

Например, одним из самых распространённых способов атомной структуры молекулы является создание модели молекулы в виде молекулярного графа. Молекулярный граф представляет собой связный неориентированный граф, находящийся во взаимно-однозначном соответствии со структурной формулой химического соединения таким образом, что вершинам графа соответствуют атомы молекулы, а рёбрам графа – химические связи между этими атомами. Понятие «молекулярный граф» является базовым для химии.

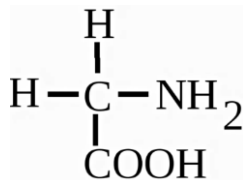


Рисунок 1 – Молекулярный граф молекулы «Глицин»

На основе молекулярного графа осуществляется описание атомно-молекулярной системы (АМС) молекулы.

Для описания атомно-молекулярных систем существует несколько способов описания координат, каждый из которых имеет свои преимущества и недостатки [8]:

1) декартовы координаты. Декартовы координаты – это пространственные координаты в выбранной пользователем декартовой системе координат, дополненные информацией о типе атома (типе химического элемента);

2) внутренние координаты;

3) координаты Z-матрицы. Z-матрица – это описание атомных координат в виде n записей (для каждого атома i в данной n-атомной системе);

4) дробные координаты.

**Описание метода.** Для данной работы было принято решение использовать описание АМС способом Z-матрицы. Он является наиболее распространённым и описывает молекулу с помощью координат молекулярной n-атомной системы (рис. 2). В качестве недостатка описания молекулярной структуры таким способом можно выделить различие в формах записи для разных программ.

Основные компоненты Z-матрицы (для каждого i атома в данной n-атомной системе):

$Z_i, Z_a, Z_b, Z_c$  – тип атома (имя химического элемента);

$i, a, b, c$  – порядковые номера;

$R_{z_i z_a}$  – межатомное расстояние между атомами  $Z_i$  и  $Z_a$ ;

$\alpha_{z_i z_a z_b}$  – валентный угол, образуемый между атомами  $Z_i$  и  $Z_a, Z_b$ ;

$\varphi_{z_i z_a z_b z_c}$  – двугранный угол, образованный атомом  $Z_i$  и плоскостью  $Z_a, Z_b, Z_c$ .

Каждому атому присваивается порядковый номер в данной n-атомной системе. Порядок описания атома i следующий:

$$\begin{cases} i > a \\ i > b. \\ i > c \end{cases} \quad (1)$$

Как видно из системы, порядковый номер «i» должен быть больше, чем номера атомов, используемых при построении.

Рассмотрим различные формы описания атомно-молекулярных систем в декартовых координатах в существующих программных комплексах:

1. Структура Z-матрицы в программе Gaussian [9].

Для того чтобы описать АМС любой молекулы, нужно задать её Z-матрицу. Например, для молекулы глицина в программе Gaussian, она будет выглядеть следующим образом:

```
C1
N2 C1 1.478822
C3 C1 1.521151 N2 110.304468
O4 C3 1.349162 C1 122.356739 N2 44.893266
O5 C3 1.213291 C1 126.358351 N2 -136.899826
H6 C1 1.111253 C2 113.633806 C3 -123.284684
H7 C1 1.109335 C2 108.691187 C3 118.707537
H8 N2 0.997497 C1 109.293714 C3 -168.846241
H9 N2 0.998174 C1 109.779802 C3 71.284892
H10 O4 0.949798 C3 110.655947 C1 -4.353087
```

Рисунок 2 – Z-матрица глицина в программе Gaussian

Таким образом, структурная формула описания АМС Z-матрицы программы Gaussian имеет вид:

$$Z_i i \ Z_a a \ R_{z_i z_a} \ Z_b b \ \alpha_{z_i z_a z_b} \ Z_c c \ \varphi_{z_i z_a z_b z_c} \quad (2)$$

2. Структура Z-матрицы в программе МОРАС [10].

Для того чтобы описать АМС любой молекулы, нужно задать её Z-матрицу. Например, для молекулы глицина в программе МОРАС она будет выглядеть следующим образом:

АТОМ NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS) NA:I	BOND ANGLE (DEGREES) NB:NA:I	TWIST ANGLE (DEGREES) NC:NB:NA:I	NA	NB	NC
1	C	0.000000	0.000000	0.000000			
2	N	1.478822	* 0.000000	0.000000	1		
3	C	1.521151	* 110.304468	* 0.000000	1	2	
4	O	1.349162	* 122.356739	* 44.893266	* 3	1	2
5	O	1.213291	* 126.358351	* -136.899826	* 3	1	2
6	H	1.111253	* 113.633806	* -123.284684	* 1	2	3
7	H	1.109335	* 108.691187	* 118.707537	* 1	2	3
8	H	0.997497	* 109.293714	* -168.846241	* 2	1	3
9	H	0.998174	* 109.779802	* 71.284892	* 2	1	3
10	H	0.949798	* 110.655947	* -4.353087	* 4	3	1

Рисунок 3 – Z-матрица ГЛИЦИНА в программе МОРАС

Таким образом, структурная формула ОАМС Z-матрицы программы МОРАС имеет вид:

$$i \ Z_i \ R_{z_i z_a} \ \alpha_{z_i z_a z_b} \ \varphi_{z_i z_a z_b z_c} \ a \ b \ c \quad (3)$$

3. Структура Z-матрицы в программе GAMESS [11].

Для того чтобы описать АМС любой молекулы, нужно задать её Z-матрицу. Например, для молекулы глицина в программе GAMESS она будет выглядеть:

```

C1
C2      1      1.5148310
O3      2      1.3546590      1      115.3041040
O4      2      1.2159740      1      128.6112280      3      178.2055680
N5      1      1.4780630      2      110.0320490      3      67.3430960
H6      1      1.1090210      2      109.3449650      3      -173.9632980
H7      1      1.1118310      2      108.6822480      3      -58.0395420
H8      3      0.9521880      2      109.9652470      1      -178.6727860
H9      5      0.9979500      1      110.2415640      2      -65.8273390
H10     5      0.9974120      1      108.9807090      2      174.0592730

```

Рисунок 4 – Z-матрица глицина в программе GAMESS

Таким образом, структурная формула описания АМС Z-матрицы программы GAMESS имеет вид:

$$Z_i \ i \ a \ R_{z_i z_a} \ b \ \alpha_{z_i z_b} \ c \ \varphi_{z_i z_c} \quad (4)$$

**Алгоритм.** Анализ различных форм описания Z-матриц позволит сделать вывод о необходимости создания алгоритма и метода трансформации записи атомно-молекулярных систем в различных вариациях.

Метод трансформации записи атомно-молекулярных систем условно делится на шесть основных этапов:

- 1) сбор данных;
- 2) преобразование;
- 3) анализ данных;
- 4) принятие решений и выполнение необходимых работ по результатам принятых решений;
- 5) обратная связь;
- 6) хранение.

**Первый этап. Сбор данных.**

Данные для трансформации Z-матрицы поступают в различных формах записи и из разных источников (файлы различного расширения): как внутренних (из базы данных), так и внешних.

**Второй этап. Преобразование.**

Таким образом, входные данные требуется конвертировать в нужный исследователю вид. Это может быть и тот же самый вид Z-матрицы, но тогда алгоритм проверяет её на правильность составления.

Алгоритм «Парсинг» файла Z-матрицы расписан по следующим шагам:

1. Начало.
2. На вход поступает файл с расширением (.out, .GJF, .res.).
3. Определяем записи по расширению поступившего файла:
  - 3.1. Если файл .out – это MORAC.
  - 3.2. Если файл .GJF – это Gaussian.
  - 3.3. Если файл .res – это GAMESS.
4. Анализ данных:
  - 4.1. Если файл расширения .out или .GJF, то:
    - 4.1.1. собираем данные описания атомно-молекулярной системы, добавляя их в двумерный массив;
    - 4.1.2. преобразуем данные массива, переставляя основные компоненты Z-матрицы к виду ОАМС программы GAMESS, согласно формуле (4).
  - 4.2. Если файл Z-матрицы имеет вид программы GAMESS, то переходим к 4 этапу (четвертый этап – принятие решений и выполнение необходимых работ по результатам принятых решений).
  - 4.3. Конец.

**Третий этап. Анализ данных.**

На этом этапе с использованием программы GAMESS происходит расчет устойчивой структуры. Программа GAMESS требует для своей работы файлы Z-матрицы, которые обычно имеют расширение \*.txt. Файл состоит из нескольких частей: программной части и части с описанием Z-матрицы. Структура входного файла, подготовленного для расчета молекулы, приведена на рисунке 6.

```

$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE MAXIT=100 COORD=ZMT NZVAR=0 ICHARG=0 $END
$CONTRL MULT=1 $END
!$SCF DIIS=.TRUE. $END
$SYSTEM TIMLIM=120 MEMORY=1000000 $END
$STATPT OPTTOL=1.0E-4 NSTEP=1800 $END
$BASIS GBASIS=PM3 $END
$GUESS GUESS=HUCKEL $END
$SCRF DIELEC=80 RADIUS=10 $END
$DATA
Structure: Glicin // Method PM3
C1
*****
*****
$END

```

Рисунок 5 – Структура входного файла, подготовленного для расчета молекулы

Алгоритм «Анализ данных» имеет следующее описание:

1. Начало.
2. На вход поступает файл расширения \*.txt.
3. Создается пакетный файл с расширением .bat, в котором содержатся несколько команд для работы с программой GAMESS. Структура .bat файла представлена на рисунке 6.

```

@echo off
cd C:/      .txt
g.bat

```

```

copy %1 input
@echo off
rem cd C:\games
if exist output erase output
if exist output erase output
if exist fockder erase fockder
if exist ciints erase ciints
rem if exist civectr erase civectr
if exist punch erase punch
if exist aoints erase aoints
if exist moints erase moints
if exist work15 erase work15
if exist work16 erase work16
if exist drtfile erase drtfile
if exist dictnry erase dictnry
@echo The computation is in progress!
@echo input file %1
C:\games\games.exe >output 2>%1.res
if exist %1.res erase %1.res
ren output %1.res

```

Рисунок 6 – Пакетный файл с расширением .bat

4. Запускает только что созданный пакетный файл.
5. Осуществляет расчет Z-матрицы, используя программу GAMESS, и выдается обратная связь в виде OUTPUT файл с расширением .res.
6. Анализирует выходной файл программа GAMESS расширения \*.res на наличие устойчивой структуры:

6.1. Если расчет прошел успешно, то из выходного документа выбираются данные оптимизированной Z-матрицы.

Оптимизация молекулярной геометрии [6] заключается в минимизации полной энергии молекулы  $E_{tot}$  при вариации координат атомов. Поскольку зависимость  $E_{tot}$  от координат ядер является поверхностью потенциальной энергии (ППЭ), оптимизация геометрии является поиском точек локальных минимумов.

6.2. Если расчет не прошел, то в созданном ОАМС содержатся ошибки.

7. Конец.

**Четвертый этап. Обратная связь.**

Из выходного документа формируется оптимизированная Z-матрица и на основе данных образуется двумерный массив. Далее компоненты массива сортируются к виду программы МОРАС согласно формуле (3) или Gaussian согласно формуле (2).

**Пятый этап. Хранение.**

Оптимизированная Z-матрица будет храниться вместе со всеми данными расчета и добавлена в базу знаний.

Блок-схема процесса работы программы представлена на рисунке 7.

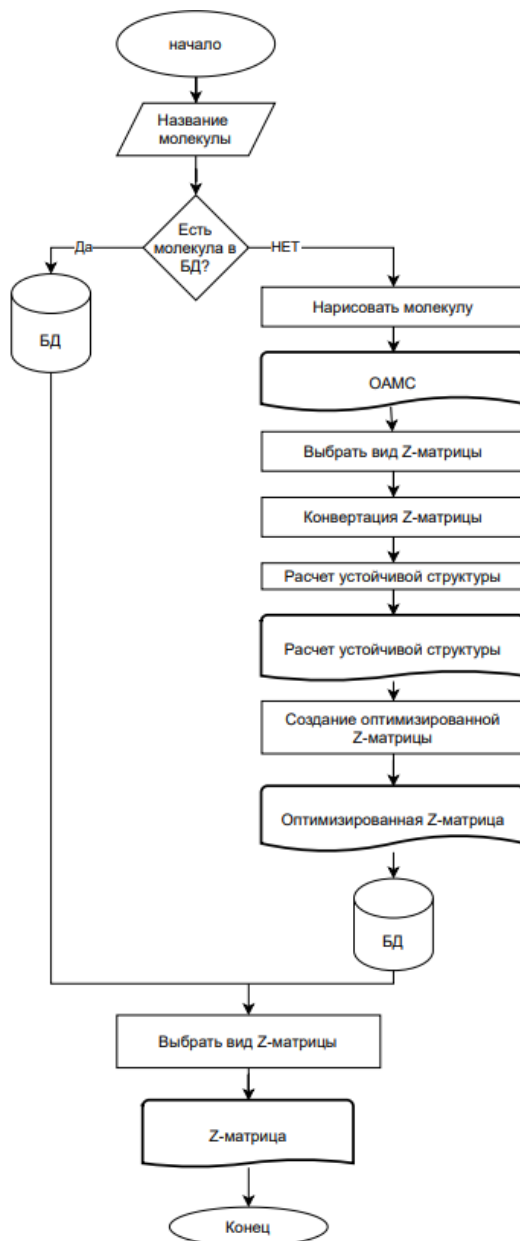


Рисунок 7 – Блок-схема алгоритма

**Заключение.** Рассмотренный подход позволяет повысить эффективность в создании записей атомно-молекулярных систем, проведении сравнительных и научных квантово-химических вычислений, сокращает время на определение устойчивой структуры молекулярных компонентов и получение готовой Z-матрицы. При создании метода используются специальные программы и дополнительные вычисления. В результате исследования был разработан специальный метод и алгоритмическое обеспечение для трансформации Z-матриц, а также получено свидетельство о регистрации программы для ЭВМ [12].

**Библиографический список**

1. Аликберова, Л. Ю. Основы строения вещества : методическое пособие / Л. Ю. Аликберова, Е. В. Савинкина, М. Н. Давыдова. – Москва : МИТХТ, 2004. – Режим доступа: <http://alhimik.ru/stroenie>, свободный. – Заглавие с экрана. – Яз. рус.

2. Тараскин, Д. В. Структура программного обеспечения для выявления потенциальных активных центров между двумя молекулами / Д. В. Тараскин, Л. И. Жарких // Вестник Технологического университета. – 2019. – Т. 22, № 12. – С. 117–121.
3. Золотарева, Н. В. Основы квантовой механики в вопросах и задачах. Модельные примеры квантовой химии : учебно-методическое пособие для студентов, обучающихся по направлениям 04.05.01 – Фундаментальная и прикладная химия, 44.03.05 – Педагогическое образование, 04.03.01 – Химия. – Астрахань, 2020.
4. Жуков, С. С. Методические указания к лабораторным работам. Часть 1 «Основы квантово-химического анализа» / С. С. Жуков. – Волгоград 2017
5. Золотарева, Н. В. Численные методы анализа в химии : учебно-методическое пособие для студентов, обучающихся по химическим направлениям и педагогическим направлениям с двумя профилями подготовки очной и очно-заочной форм обучения / Н. В. Золотарева. – Астрахань, 2020.
6. Демидович, Н. А. Автоматизация обработки квантово-химических вычислений с использованием GAMESS / Н. А. Демидович // XXVI Региональная конференция молодых ученых и исследователей Волгоградской области : сборник материалов конференции. – Волгоград, 2022. – С. 165–167.
7. Budzelaar, P. H. M. Geometry optimization using generalized, chemically meaningful constraints / P. H. M. Budzelaar // Journal of Computational Chemistry. – 2007. – Т. 28, № 13. – С. 2226–2236.
8. Игнатов, С. К. Квантово-химическое моделирование атомно-молекулярных процессов : учебное пособие / С. К. Игнатов. – Нижний Новгород 2019. – 79 с.
9. Программа Gaussian. – Режим доступа: <https://gaussian.com>, свободный. – Заглавие с экрана. – Яз. рус.
10. Программа MOPAC® – Режим доступа: <http://openmopac.net>, свободный. – Заглавие с экрана. – Яз. рус.
11. Программа Gamess. The General Atomic and Molecular Electronic Structure System (GAMESS) is a general ab initio quantum chemistry package. – Режим доступа: [www.msg.chem.iastate.edu/gamess/index.html](http://www.msg.chem.iastate.edu/gamess/index.html), свободный. – Заглавие с экрана. – Яз. рус.
12. Программа для ЭВМ «TFinG» № 2022614450 от 22.03.22.

#### References

1. Alikberova, L. Yu., Savinkina, E. V., Davydova, M. N. *Osnovy stroeniya veshchestva : metodicheskoe posobie* [Fundamentals of substance structure : tutorial]. Moscow, 2004. Available at: <http://alhimik.ru/stroenie>.
2. Taraskin, D. V., Zharkikh, L. I. *Struktura programmnogo obespecheniya dlya vyyavleniya potentsialnykh aktivnykh tse ntrov mezhdu dvumya molekulami* [Software structure for identifying potential active centers between two molecules]. *Vestnik Tekhnologicheskogo universiteta* [Vestnik Technological University], 2019, vol. 22, no. 12, pp. 117–121.
3. Zolotareva, N. V. *Osnovy kvantovoy mekhaniki v voprosakh i zadachakh. Modelnye primery kvantovoy khimii : uchebno-metodicheskoe posobie dlya studentov, obuchayushchikhsya po napravleniyam 04.05.01 – Fundamentalnaya i prikladnaya khimiya, 44.03.05 – Pedagogicheskoe obrazovanie, 04.03.01 – Khimiya* [Fundamentals of Quantum Mechanics in questions and problems. model examples of quantum chemistry : textbook for students studying in the areas of 04.05.01 – Fundamental and Applied Chemistry, 44.03.05 – Pedagogical Education, 04.03.01 – Chemistry]. Astrakhan, 2020.
4. Zhukov, S. S. *Metodicheskie ukazaniya k laboratornym rabotam. Chast 1 «Osnovy kvantovo-khimicheskogo analiza»* [Methodological instructions for laboratory works. Part 1 "Fundamentals of quantum-chemical analysis"]. Volgograd, 2017.
5. Zolotareva, N. V. *Chislennyye metody analiza v khimii : uchebno-metodicheskoe posobie dlya studentov, obuchayushchikhsya po khimicheskim napravleniyam i pedagogicheskim napravleniyam s dvumya profilyami podgotovki ochnoi i ochno-zaochnoy form obucheniya* [Numerical methods of analysis in chemistry : textbook for students studying in chemical directions and pedagogical directions with two training profiles of full-time and part-time forms of training]. Astrakhan, 2020.
6. Demidovich, N. A. *Avtomatizatsiya obrabotki kvantovo-khimicheskikh vychisleniy s ispolzovaniem GAMESS* [Automation of quantum-chemical calculations processing using GAMESS]. *XXVI Regionalnaya konferentsiya molodykh uchenykh i issledovateley Volgogradskoy oblasti : sbornik materialov konferentsii* [XXVI Regional Conference of young scientists and researchers of Volgograd region : collection of conference materials]. Volgograd, 2022, pp. 165–167.
7. Budzelaar, P. H. M. Geometry optimization using generalized, chemically meaningful constraints. *Journal of Computational Chemistry*, 2007, vol. 28, no. 13, pp. 2226–2236.
8. Ignatov, S. K. *Kvantovo-khimicheskoe modelirovanie atomno-molekulyarnykh protsessov : uchebnoe posobie* [Quantum-chemical modeling of atomic-molecular processes : textbook]. Nizhniy Novgorod, 2019. 79 p.
9. *Programma Gaussian* [Gaussian program]. Available at: <https://gaussian.com>.
10. *Programma MOPAC®* [MOPAC® program]. Available at: <http://openmopac.net>.
11. *Gamess program. The General Atomic and Molecular Electronic Structure System (GAMESS) is a general ab initio quantum chemistry package*. Available at: [www.msg.chem.iastate.edu/gamess/index.html](http://www.msg.chem.iastate.edu/gamess/index.html).
12. *Programma dlya EVM «TFinG»* [Computer program "TFinG"], no. 2022614450 of 22.03.22.