

РЕАЛИЗАЦИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПОИСКА СХОЖИХ ПРИЗНАКОВ И УСТАНОВЛЕНИЕ АКТИВНЫХ ЦЕНТРОВ В СТРУКТУРЕ ВЗАИМОДЕЙСТВИЯ БЕЛКОВ С МАЛЫМИ МОЛЕКУЛАМИ SO_2 , H_2S , NO_2 , NH_3

IMPLEMENTATION OF MATHEMATICAL MODELS FOR THE SEARCH OF SIMILAR CHARACTERISTICS AND THE ESTABLISHMENT OF ACTIVE CENTERS IN THE STRUCTURE OF THE INTERACTION OF PROTEINS WITH SMALL MOLECULES SO_2 , H_2S , NO_2 , NH_3

**N. Zolotareva
E. Evsina**

Summary. It offers the implementation of a mathematical model for the search of similar characteristics and the establishment of active centers in the structure of the interaction of proteins with small molecules SO_2 , H_2S , NO_2 , NH_3 . The automated system implements several main tasks, each of which performs a number of functions: 1) simplification of work with files during the whole process of modeling of intermolecular interactions; 2) provision of information about active centers in protein structures and targets of small molecules; 3) graphical support in interpreting the results, which simplifies the analysis of numerical data. The automated system developed in the framework of this work is integrated with the database. As DBMS we used non-commercial software FireBird 3.0, freely distributed on the Internet, FireBird 3.0 runs on many operating systems, including Windows. The use of quantum chemical apparatus in solving chemical and environmental problems can provide a serious mathematical basis for the formation of calculations, as the mechanisms of the processes, and the basis for the creation of antidotes to various toxic substances. Considered the stages of formation of models of interaction on the example of small molecules, the algorithm of calculation of the main indicators and criteria for the selection of optimal configurations can be used to solve environmental and chemical challenges and implemented into the system of environmental risk analysis.

Keywords: mathematical model, database, interaction models, chemical and environmental problems, calculation algorithm.

Золотарева Наталья Валерьевна
К.т.н., доцент, ФГБОУ ВО «Астраханский
государственный университет», Астрахань
zoloto.chem@mail.ru

Евсина Елена Михайловна
К.т.н., доцент, ГАОУ АО ВО «Астраханский
государственный архитектурно-строительный
университет», Астрахань
evsinalena@mail.ru

Аннотация. Предлагается реализация математической модели поиска схожих признаков и установление активных центров в структуре взаимодействия белков с малыми молекулами SO_2 , H_2S , NO_2 , NH_3 . Автоматизированная система реализует несколько основных задач, каждая из которых выполняет ряд функций: 1) упрощение работы с файлами на протяжении всего процесса моделирования межмолекулярных взаимодействий; 2) предоставление информации об активных центрах в белковых структурах и о мишенях воздействия малых молекул; 3) графическое сопровождение при интерпретации результатов, что упрощает анализ числовых данных. Разработанная в рамках данной работы автоматизированная система интегрирована с базой данных. В качестве СУБД использовалось некоммерческое программное обеспечение FireBird 3.0, свободно распространяемое в сети Интернет, FireBird 3.0 работает под управлением многих операционных систем, в том числе Windows. Применение квантово-химического аппарата в решении химико-экологических задач способно предоставить серьезную математическую базу расчетов для формирования, как механизмов происходящих процессов, так и базу для создания антидотов к различным токсическим веществам. Рассмотренные этапы формирования моделей взаимодействия на примере малых молекул, алгоритм расчета основных показателей и критерии отбора оптимальных конфигураций могут быть использованы для решения химико-экологических задач и внедрены в систему анализа экологического риска.

Ключевые слова: математическая модель, база данных, модели взаимодействия, химико-экологические задачи, алгоритм расчета.

Введение

В настоящее время, одним из перспективных и актуальных направлений развития науки и техники является разработка технологий, средств и методов мониторинга и прогнозирования состояния объектов окружающей среды, а также предотвращения и ликвидации загрязнений. Поэтому, вопросы экологической безопасности и мониторинговые исследования по надзору за уровнем загрязнения как природного, так и техногенного характера в окружающей среде выходят на первый план. Мониторинг служит целому ряду целей, от контроля соответствия химического и экологического статуса до защиты конкретных объектов окружающей среды. При этом, существующие различные способы отбора проб, *in vitro* и био- и химико-аналитические методы постоянно совершенствуются для улучшения мониторинга загрязнения. Основные улучшения направлены на включение методов скрининга и идентификации более широкого спектра химических веществ, высокую чувствительность обнаружения токсикантов, разработку стандартизированных протоколов для химических, токсикологических и экологических оценок. Однако, ни один метод или комбинация методов не могут удовлетворить различные цели мониторинга. Современные подходы к мониторингу имеют тенденцию подчеркивать либо целевое воздействие, либо обнаружение эффекта. К примеру, химико-аналитический мониторинг за состоянием объектов окружающей среды позволяет установить общую тенденцию к накоплению токсичных веществ [1, 2, 3]. При этом, основное требование заключается в корректном воссоздании необходимых условий проведения лабораторного эксперимента, поэтому, зачастую количественная оценка на основании только одного метода далеко не всегда оказывается корректной [4, 5, 6].

Необходимо отметить, что, независимо от поставленной цели, оценка результатов мониторинга будет эффективной и существенно выиграет от получения и связывания информации не только о наличии химических веществ, относящихся к определенному классу опасности, но, также о потенциально неблагоприятных эффектах на уровне локализации мишеней воздействия.

Цель исследования

Моделирование, в таком случае, выполняет ключевую роль при прогнозировании поведения токсикантов в объектах окружающей среды, при оценке экологического риска, а также, при установлении механизмов протекания процессов в биосистемах [7, 8, 9]. Поскольку в процессе моделирования обрабатывается большой массив данных, а также часть информации подлежит хранению и их дальнейшей интерпретации возникла по-

требность в разработке автоматизированной системы управления массивами данных [10, 11].

Материал и методы исследования

Автоматизированная система направлена на реализацию основных задач:

1. Выполнение автоматизированной работы с входными и выходными документами на протяжении всего процесса моделирования. Ключевые функции: а) формирование исходной информации о геометрии активного центра белковой системы, аминокислотной последовательности; б) генерация материалов структурных файлов для последующей обработки программными квантово-химическими комплексами MOPAC2016 [12] и Gamess [13]; в) обработка, систематизация оптимизированных результатов вычислений программных комплексов и внесение их в базу данных;

2. Вывод информации о составе и структурных изменениях активных центров локализованных участков в белковых системах и о мишенях воздействия токсичных веществ, на примере малых молекул. Ключевые функции: а) вычисление термодинамических параметров, образующихся при взаимодействии межмолекулярных систем; б) многокритериальный поиск по базе данных.

3. Интерпретация результатов вычислений с использованием визуального сопровождения, что позволяет упростить анализ массива данных. Ключевые функции: а) визуализация активных центров локализованных участков в белковых системах; б) вывод молекулярной диаграммы аминокислотной последовательности (часто встречающаяся в белковой системе) в непосредственной близости к токсичному соединению, малой молекуле.

Разработанная в рамках данной работы автоматизированная система интегрирована с базой данных. В качестве СУБД использовалось некоммерческое программное обеспечение FireBird 3.0 свободно распространяемое в сети Интернет и работающее под управлением многих операционных систем [14]. На рис. 1а и 1б приведены схемы связей данных.

В базу данных внесена информация о структурных и энергетических параметрах малых молекул SO₂, H₂S, NO₂, NH₃, CO (*structure_one*). Заполнение таблицы осуществляется после успешного завершения вычислительного итерационного процесса, в случае если найдена оптимальная конфигурация молекулы. Выборка данных производится из выходных файлов формата «*.txt». В таблице 1 представлена характеристика основных параметров.

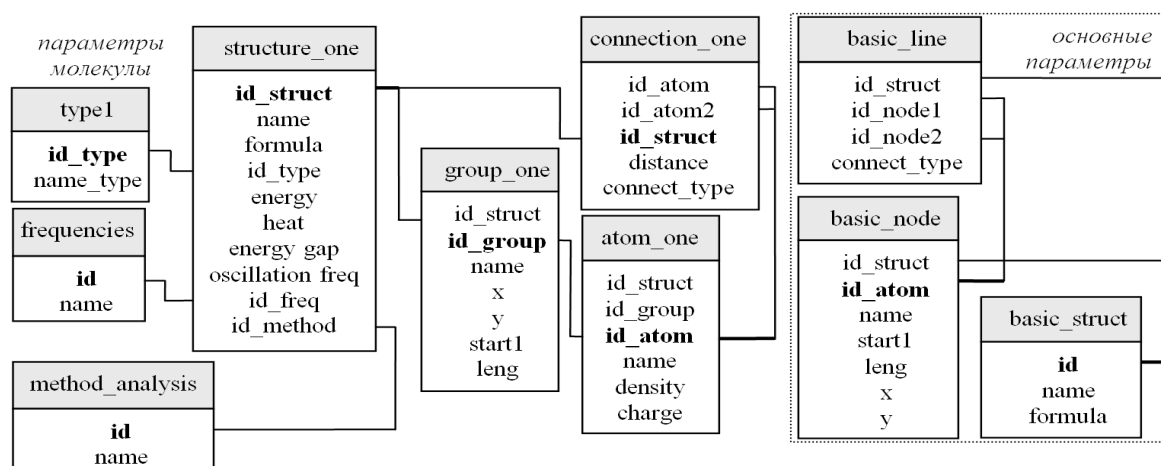


Рис. 1а. Схема связей данных

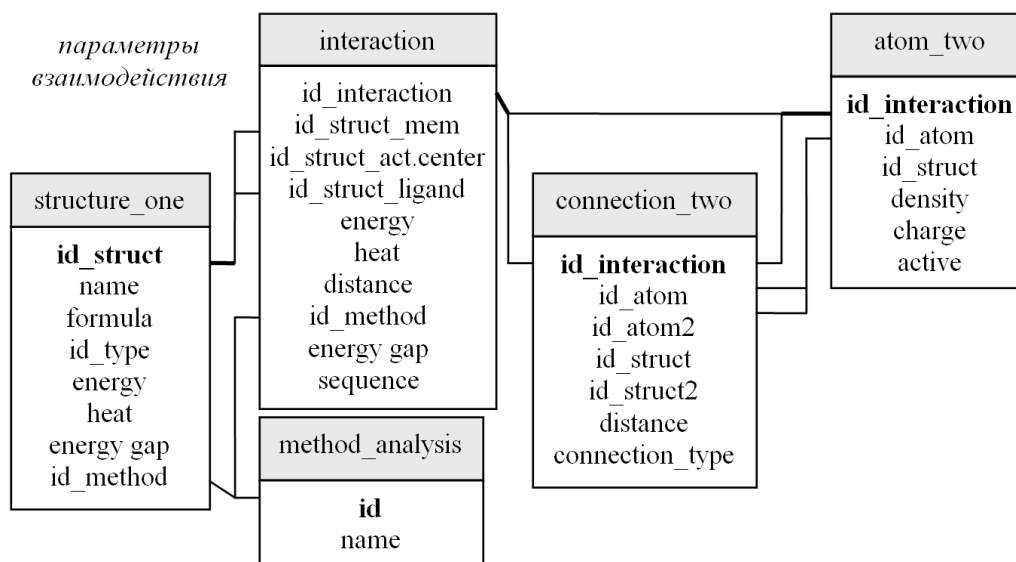


Рис. 1б. Схема связей данных взаимодействия

Однако, если на этапе итерационного процесса, не достигнуто оптимальное положение структуры, то вычислительный процесс запускается повторно с новыми координатами.

Для характеристики группы атомов в активных центрах белковых систем введены обозначения, фиксирующие взаимное расположение в системе координат x, y и положение нижних индексов в функциональных группах молекул. В таблице 2 представлена характеристика основных параметров.

При формировании схемы «молекулярная диаграмма» ключевыми параметрами являются величины зарядов

q_i (по Малликену) и значения электронной плотности ρ_i на атомах в аминокислотной последовательности, поэтому они также вносятся в базу данных. В таблице 3 представлены идентификаторы как по структуре, по функциональным группам атомов, так и по типу атомов.

Поскольку атомы объединены в функциональные группы, соответственно, необходимо учитывать, что разные типы атомов по-разному могут быть связаны друг с другом, количественно это характеризуется степенью связывания. В таблице 4 приведены основные параметры, которые используются для описания химических связей между атомами.

Таблица 1. Structure_one

id_struct	Идентификатор структуры
name	Наименование (150 знаков ввода)
formula	Формула (150 знаков ввода)
id_type	Тип
energy	Общая энергия (число двойной точности)
energy	Общая энергия (число двойной точности)
heat	Теплота образования (число двойной точности)
energy gap	Энергетическая щель (число двойной точности)
oscillation frequencies	Частоты колебаний (число двойной точности)
id_freq	Идентификатор частоты колебаний
id_method	Идентификатор метода вычислений

Таблица 2. Group_one

id_struct	Идентификатор структуры
id_group	Идентификатор группы атомов
name	Обозначение группы атомов (7 знаков ввода)
X	Положение на схеме (X-координата)
Y	Положение на схеме (Y-координата)
start1	Положение нижнего индекса
leng	Длина нижнего индекса

Таблица 3. Atom_one

id_struct	Идентификатор структуры
id_group	Идентификатор группы атомов
id_atom	Идентификатор атома
name	Обозначение атома (7 знаков ввода)
density	Электронная плотность (число двойной точности)
charge	Заряд (число двойной точности)

Таблица 4. Connection_one

id_atom	Идентификатор атома
id_atom2	Идентификатор атома 2
id_struct	Идентификатор структуры
distance	Длина связи (число двойной точности)
connect_type	Одинарная, двойная, тройная связи

Результаты исследования и их обсуждение

В структуру базы данных дополнительно включено три справочника:

- ♦ **type1** содержит сведения о типах малых молекул (SO₂, H₂S, NO₂, NH₃ и др.). Для ввода названия соединения в поле name_type предусмотрено 25 знаков;

- ♦ **method_analysis** содержит сведения о вычислительных методах (semi empirical: PM3, PM7 и другие /ab initio: RHF, MP2 и другие) и наборы базисных функций, используемых в расчетах. Для ввода данных в текстовое поле name предусмотрено 10 знаков;

- ♦ **frequencies** содержит экспериментальные и вычисленные сведения ИК-частот колебаний малых молекул.

Визуальное сопровождение в виде схемы «молекулярная диаграмма» осуществляется в пользовательском окне. База данных также содержит таблицы с информацией об основных структурах (`basic_struct`), об основных узлах (`basic_node`) и о длинах химических связей (`basic_line`). Таблица `Basic_node` содержит идентификатор по структуре и по атомам; добавлены числовые поля с нижним индексом (`start1`); координаты узла X и Y фиксируют месторасположение атомов на молекулярной диаграмме.

Параметры химических связей основных структур, которые классифицированы по кратности, представлены в таблице `Basic_line`. Все поля определены целыми числовыми данными. Например, если в структуре имеется одинарная связь (σ), это соответствует единице, двойной связи (σ, π) соответствует 2, если присутствует тройная связь ($\sigma, 2\pi$) между атомами, то поле определено тройкой. Необходимо отметить, что наличие начальной и конечной точек связывания в таблице представлено полями `id_node1`, `id_node2`.

В результате, первый модуль отвечает за автоматизацию информации, хранящийся в таблицах базы данных и позволяет предоставить пользователю выбор показателей, которые могут быть вынесены на молекулярную диаграмму. Таким образом, можно проводить сравнительный анализ между отдельными фрагментами белковой системы.

Входными данными первого модуля являются:

1. метод и базис расчета, параметры оптимизации;
2. название исследуемой молекулы, принадлежность к конкретному типу фрагмента в белковой системе;
3. начальные геометрические параметры (декартовы координаты, z -матрица во внутренних координатах) малой молекулы;
4. энергетические, силовые и зарядовые параметры малой молекулы.

Выходными данными первого модуля являются:

- ♦ энергетические параметры молекулы (общая энергия, энергетическая щель);
- ♦ таблица оптимизированных геометрических параметров молекулы (длина химической связи);
- ♦ зарядовые и силовые характеристики молекулы;
- ♦ двумерная схема молекулярной модели («молекулярная диаграмма»);
- ♦ данные входных параметров второго модуля в режиме моделирования межмолекулярных взаимодействий и расчета основных параметров, образующихся систем.

Идентификация моделей взаимодействия между фрагментом белковой системы и малой молекулой обеспечивается полем `id_interaction`. Таблица данных

содержит идентификатор по структуре компонента белка (`id_struct_mem`), по структуре активного центра (`id_struct_active center`) и по структуре малой молекулы (`id_struct_ligand`). Аналогично однокомпонентным структурам, поле `id_method` содержит информацию о методе и базисе расчета. Выбор метода расчета определяется адекватностью описания геометрии молекул и воспроизведению частот колебаний при сопоставлении с экспериментальными параметрами.

Необходимо отметить, что в структуре активного центра между отдельными функциональными группами аминокислот и малой молекулой формируются водородные связи, в диапазоне 1,70...2,5 Å. Таким образом, межъядерное расстояние является первым критерием, по которому осуществляется выборка и внесение сведений об активных центрах в базу данных. Тип данных поля `distance` числовой, а размер определен двойной с плавающей точкой. Величина общей энергии (кДж/моль) характеризует систему в целом, а энергетическая щель (эВ) характеризует мягкость/жесткость образующейся системы. Тип данных поля `energy`, `heat` и `energy gap` определены числом, размер поля соответствует двойной с плавающей точкой. Аминокислотная последовательность (`sequence`) генерируется из исходного файла белковой системы и позволяет оценить схожий набор аминокислот, расположенных в непосредственной близости от малой молекулы и спрогнозировать поведение малых молекул при взаимодействии с другими видами белков.

Основные параметры электронной плотности и заряды на атомах в контактной системе вносятся в таблицу `atom_two`. Идентификация осуществляется по типу взаимодействия (`id_interaction`); по атому (`id_atom`) и структуре (`id_struct`). Тип данных полей задается целым числом, а размер полей `density` и `charge` определен двойной с плавающей точкой. Степень смещения электронной плотности в структуре от одного атома к другому имеет дробное значение. Идентификатор активности атомов (`active`) позволяет установить реакционную способность атомов, участвующих во взаимодействии. Величина «нуль» в поле `active` свидетельствует о наличии слабой активности атома; единица классифицирует повышенную активность атомов.

Таким образом, второй модуль автоматизированной системы обеспечивает управление параметрами во взаимодействующей системе. В результате осуществляется отбор наиболее стабильных конфигураций, устанавливаются активные центры на поверхности белковых систем и формируются модельные схемы взаимодействий.

Входными данными модуля являются:

- ♦ начальные геометрические параметры (декартовы координаты, z -матрица во внутренних координатах) взаимодействующей системы;

- ◆ энергетические и зарядовые характеристики системы;
- ◆ метод и базис расчета, параметры оптимизации;
- ◆ название и формула соединения;
- ◆ расстояние между атомами взаимодействующих молекул.

Выходными данными модуля являются:

1. оптимизированные геометрические параметры структуры активного центра;
2. минимизированные энергетические параметры структуры активного центра;
3. зарядовые характеристики атомов в структуре активного центра;
4. активность атомов в локализованном центре и аминокислотная последовательность;
5. двухмерная модель взаимодействия.

Выводы или заключение

Полученные результаты позволяют сформировать обобщенный взгляд на тематику проблемы. Применение квантово-химического аппарата в решении химико-экологических задач способно предоставить серьезную математическую базу расчетов для формирования, как механизмов происходящих процессов, так и базу для создания антидотов к различным токсическим веществам. Рассмотренные этапы формирования моделей взаимодействия на примере малых молекул, алгоритм расчета основных показателей и критерии отбора оптимальных конфигураций могут быть использованы для решения химико-экологических задач и внедрены в систему анализа экологического риска.

ЛИТЕРАТУРА

1. Kermani M., Jonidi Jafari Ah., Rezaei Kalantari R., Sadat Sakhaet F., Sadat Kahe T., Dowlati M. Evaluation of chronic obstructive pulmonary disease attributed to atmospheric O₃, NO₂ and SO₂ in Tehran city, from 2005 to 2014. *Iranian Journal of Health, Safety & Environment*, 2014. vol. 4. no. 3. P. 758–766.
2. Hazardous Substances Data Bank [Internet]. National Library of Medicine (US): Bethesda (MD). [Last Revision Date 2010 Jan. 21; cited 2018 Sept. 5]. Available from: <https://toxnet.nlm.nih.gov>.
3. Esehie A. Beneficial effect of a hydrogen sulphide donor (sodium sulphide) in an ovine model of burn- and smoke-induced acute lung injury. *Br. J. Pharmacol*, 2009. vol. 158. no. 6. P. 1442–1453.
4. Kolluru G. K. Hydrogen sulfide chemical biology: pathophysiological roles and detection. *Nitric Oxide*, 2013. vol.35. P. 5–20.
5. Malone Rubright S. L. Environmental toxicology of hydrogen sulfide. *Nitric Oxide*, 2017. vol. 71. P. 1–13.
6. Jiang J. Hydrogen sulfide-mechanisms of toxicity and development of an antidote. *Sci. Rep.*, 2016. no. 6. P. 20831.
7. Subba J. R., Thammakhet Ch., Thavarungkul P., Kanatharana P. Distributions of SO₂ and NO₂ in the lower atmosphere of an industrial area in Bhutan. *Journal of environmental science and health, part A*, 2016. P. 1–10.
8. Bobbitt N. Scott, Mendonca Matthew L., Ashlee J. Howarth, Timur Islamoglu, Hupp Joseph T., Omar K. Farha, Randall Q. Metal-organic frameworks for the removal of toxic industrial chemicals and chemical warfare agents. *Snurr-Chem. Soc. Rev.*, 2017. vol. 46. P. 3357–3385.
9. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. (2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235–242. [Last Revision Date 2000 Jan. 1; cited 2018 Sept. 5]. Available from: <http://www.rcsb.org/pdb/home/home.do>.
10. Литвин Д. Б., Дроздова Е. А. Математическое моделирование в среде визуального программирования // *Современные наукоемкие технологии*. — 2013. — № 6; URL: <http://top-technologies.ru/ru/article/view?id=31989> (дата обращения: 05.09.2018).
11. Дваденко М. В., Привалова Н. М., Кудяева И. Ю., Степура А. Г. Выбор адсорбента для очистки сточных вод // *Современные наукоемкие технологии*. 2010. — № 10.; URL: <https://expeducation.ru/ru/article/view?id=11244> (дата обращения: 05.09.2018).
12. MOPAC2016, James J. P. Stewart, *Stewart Computational Chemistry*, Colorado Springs, CO, USA, <HTTP://OpenMOPAC.net> (2016).
13. M. S. Gordon and M. W. Schmidt. *Advances in Electronic Structure Theory: GAMESS a Decade Later, Theory and Applications of Computational Chemistry*, Ch. 41, C. E. Dykstra, G. Frenking, K. S. Kim, G. E. Scuseria, Eds., Elsevier, 2005.
14. Firebird Quick Start Guide Ch. 1 Using Firebird. IBPhoenix Editors. Firebird Project members. [Last Revision Date 2016 Apr. 26] Available from: <http://www.firebirdsql.org/en/documentation/>.

© Золотарева Наталья Валерьевна (zoloto.chem@mail.ru), Евсина Елена Михайловна (evsinalena@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»