

УДК 004.652.4, 004.043, 004.942/.001.57

**СИСТЕМА УПРАВЛЕНИЯ ДАННЫМИ ПРИ ПОИСКЕ ОПТИМАЛЬНЫХ
КОНФИГУРАЦИЙ И УСТАНОВЛЕНИЕ АКТИВНЫХ ЦЕНТРОВ**

Н.В. Золотарева, А.Ю. Макаренко

В статье рассматриваются основные принципы построения системы управления данными для поиска оптимальных конфигураций и установление активных центров (Molecular Model), описаны составные части данной автоматизированной системы, их принципы работы, функционал и структура базы данных для хранения результатов вычислений.

Ключевые слова: компьютерное моделирование, установление активных центров, поиск оптимальных конфигураций.

Key words: computer modelling, determination of active centers, search for the optimal configuration.

В химии процессов, происходящих при обычных температурах, важную роль играют межмолекулярные, а именно водородные, силы взаимодействия. Велика роль подобных взаимодействий в стабилизации конденсированных состояний молекулярных систем, например, воды, в стабилизации биополимеров, а также протекании биохимических процессов.

Данная статья является продолжением серии работ [1–3], посвященных моделированию механизмов воздействия низкомолекулярных токсичных соединений на биополимеры биологической системы, а точнее, клеточной мембраны. Поскольку клеточная мембрана представлена биомембранным слоем, состоящим из слоев липидных молекул со встроенными в них белками и углеводами [6], модель мембраны может быть представлена в виде набора структурных графов, изображающих отдельно белковую, липидную и углеводную компоненту. Но даже такие компоненты представляют собой громоздкие конструкции, состоящие из многих сотен атомов. Для проведения компьютерного моделирования из множества элементов жидкостно-мозаичной структуры нами были выделены схожие, тесно связанные друг с другом функциональные группы. В результате сложная молекулярная модель дифференцирована на небольшие составляющие, и все расчеты проводятся только для этих компонентов.

Используемый в работе программный комплекс Gamess (US)* реализует только численные методы оптимизации геометрии молекул. Дальнейшая обработка полученных результатов и их интерпретация пользователем проводится самостоятельно, что зачастую сводится к анализу большого количества вычислений. В связи с этим возникла потребность в создании автоматизированной системы управления результатами квантово-химических расчетов.

Разработанная автоматизированная система Molecular Model [5] реализует несколько основных задач, каждая из которых выполняет ряд функций.

1. Упрощение работы с файлами на протяжении всего процесса моделирования межмолекулярных взаимодействий. Основные функции:

- оформление исходных документов, содержащих информацию о первоначальной геометрии молекул. Генерация исходных текстов для дальнейшей обработки программным комплексом Gamess (US);

* Функционирование поддерживается группой профессора М. Гордона (M. Gordon, Ames Laboratory/Iowa State University, USA).

ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

- обработка оптимизированных результатов вычислений программного комплекса Gamess (US) и внесение их в базу данных.

2. Предоставление информации о потенциальных «мишенях» воздействия токсичного вещества на биологические системы:

- расчет термодинамических характеристик, образующихся при взаимодействии межмолекулярных систем;
- многокритериальный поиск по базе данных.

3. Графическое сопровождение при интерпретации результатов, что упрощает анализ числовых данных:

- визуализация исследуемых объектов с учетом индексов реакционной способности.

Автоматизированная система Molecular Model состоит из нескольких модулей:

- первый – BioMolDiagrams – позволяет на основании результатов расчета составлять молекулярные диаграммы биополимеров клеточной мембраны [4];
- второй – ModelInteractions – предназначен для формирования схем взаимодействий и установления активных центров в молекуле биополимера.

Система Molecular Model интегрирована с базой данных. В качестве СУБД использовался FireBird 2.1. Это некоммерческое программное обеспечение, свободно распространяемое в сети Интернет (www.firebirdsql.org), работает под управлением многих операционных систем, в том числе и Windows. Рассмотрим структуру БД, хранящую информации об однокомпонентных (до воздействия) и двухкомпонентных (после взаимодействия) системах.

На рис. 1 приведена схема базы данных автоматизированной системы Molecular Model.

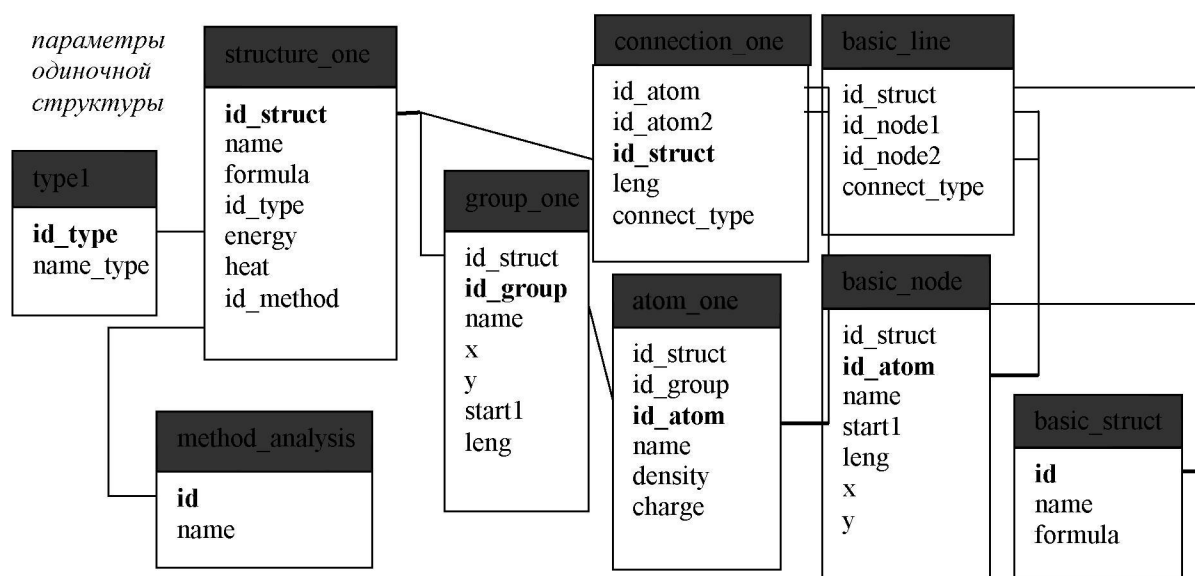


Рис. 1. Схема связей данных

Данная структура позволяет хранить информацию о следующих сущностях и их параметрах.

1. Отдельная однокомпонентная структуры (*structure_one*). База данных хранит информацию о параметрах индивидуальных молекул. Заполнение таблицы происходит после завершения процесса оптимизации молекул программой Gamess. Из выходных файлов «.txt» производится выборка тех данных, которые указаны в таблицах 1–4.

ПРИКАСПИЙСКИЙ ЖУРНАЛ:
управление и высокие технологии № 2 (10) 2010

Таблица 1

Structure_one

id_struct	INTEGER	Идентификатор структуры
name	VARCHAR(150)	Наименование
formula	VARCHAR(150)	Формула
id_type	INTEGER	Тип
energy	DOUBLE PRECISION	Общая энергия
heat	DOUBLE PRECISION	Теплота образования
id_method	INTEGER	Метод анализа

Размеры полей *name* и *formula* составляют 150 знаков ввода. Размер и тип чисел, вводимых в поля *energy* и *heat*, соответствуют двойной точности. Общая энергия и теплота образования молекулы автоматически заполняют базу данных только в том случае, если в ходе итерационного процесса была найдена оптимальная конфигурация. Окончанием процесса оптимизации в программе Gamess является показатель «Equilibrium Geometry Located». Если структура не достигла оптимального положения, то процесс запускается повторно с новыми координатами.

2. Группы атомов (*group_one*). Для характеристики группы атомов в таблице 2 *group_one* приведены условные обозначения; взаимное расположение в системе координат *x**y* и положение нижних индексов для функциональных групп молекулы.

Таблица 2

Group_one

id_struct	INTEGER	Идентификатор структуры
id_group	INTEGER	Идентификатор группы атомов
name	VARCHAR(7)	Обозначение группы атомов
X	INTEGER	Положение на чертеже (X-координата)
Y	INTEGER	Положение на чертеже (Y-координата)
start1	INTEGER	Положение нижнего индекса
leng	INTEGER	Длина нижнего индекса

Поскольку в структуре возможно повторение функциональных групп, например -CH₂-CH₂-, было принято решение об идентификации атомов по группам. Поле *name* может быть определено графическим представлением функциональной группы (-NH₂, -SO₃H). Остальные данные определяют положение группы атомов на чертеже и положение нижних индексов.

3. Отдельные атомы структуры (*atom_one*). Для формирования молекулярных диаграмм параметры зарядов на атомах *q_i* и значения электронной плотности *π_i* являются ключевыми, поэтому они также заносятся в базу данных. За условными изменениями в электронной конфигурации можно проследить по значениям зарядов, определяющих некоторое интегральное значение электронной плотности вблизи каждого атома. Размер полей *density* и *charge* соответствует значениям с двойной плавающей точкой. В таблице 3 имеются идентификаторы как по структуре, по функциональным группам атомов, так и по типу атомов.

ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Таблица 3

Atom_one

id_struct	INTEGER	Идентификатор структуры
id_group	INTEGER	Идентификатор группы атомов
id_atom	INTEGER	Идентификатор атома
name	VARCHAR(7)	Обозначение атома
density	DOUBLE PRECISION	Электронная плотность
charge	DOUBLE PRECISION	Заряд

4. Связи между отдельными атомами структуры (connection_one). Атомы объединены в группы, соответственно необходимо учитывать, что разные атомы по-разному могут быть связаны друг с другом, это количественно определяется степенью связывания. В таблице 4 приведены параметры, которые используются для описания связей между атомами.

Таблица 4

Connection

id_atom	INTEGER	Идентификатор атома
id_atom2	INTEGER	Идентификатор атома
id_struct	INTEGER	Идентификатор структуры
leng	DOUBLE PRECISION	Длина связи
connection_type	INTEGER	Одинарная, двойная, тройная связи

Поле длины связи определено двойной точностью, остальные поля таблицы классифицированы целыми числовыми значениями.

Кроме того, в структуре БД (рис. 1) представлены 2 справочника:

- type1 – содержит типы однокомпонентных структур (белок, углевод, липид и др.), табл. 5.

Таблица 5

Type1

id_type	INTEGER	Идентификатор
name_type	VARCHAR(25)	Наименование

Максимальное число знаков для ввода в поле name_type составляет 25, это позволяет вписать общее название класса, в которую включена исследуемая молекула;

- method_analysis – методы анализа структур (PM3, 6-31G* и др.), табл. 6.

Таблица 6

Method_analysis

id	INTEGER	Идентификатор
name	VARCHAR(10)	Наименование

Количество знаков в текстовом поле name составляет 10.

Для графического отображения структур (схем молекулярных диаграмм) в пользовательском окне, рассмотрим схемы связей данных, хранящихся в таблицах 7–9. База содержит таблицы с информацией о базовых структурах (basic_struct), об узлах (basic_node) и о длинах связей (basic_line).

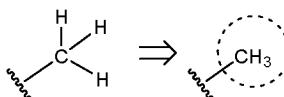
**ПРИКАСПИЙСКИЙ ЖУРНАЛ:
управление и высокие технологии № 2 (10) 2010**

Таблица 7

Basic_struct		
id	INTEGER	Идентификатор
name	VARCHAR(150)	Наименование
formula	VARCHAR(150)	Формула

Колонки *name* и *formula* определены текстовым типом данных, размер поля составляет 150 знаков.

Для отображения молекулы с учетом всех атомов, принято решение об использовании нижних индексов, которые задают количественный состав какого-либо атома в структуре, например:



В табл. 8 приведены узлы для базовых структур.

Таблица 8

Basic_node		
id_struct	INTEGER	Идентификатор структуры
id_atom	INTEGER	Идентификатор атома
name	VARCHAR(20)	Наименование
start1	INTEGER	Начало текста нижним индексом
leng	INTEGER	Длина текста нижним индексом
x	INTEGER	X-координата узла
y	INTEGER	Y-координата узла

Таблица включает идентификатор по структуре и по атомам; включены числовые поля нижнего индекса (*start1*); координаты узла *X* и *Y* указывают на месторасположение атомов на схеме, размер поля определен длинным целым числом.

В табл. 9 представлены параметры связей базовых структур, которые классифицированы по кратности.

Таблица 9

Basic_line		
id_struct	INTEGER	Идентификатор структуры
id_node1	INTEGER	Начало связи
id_node2	INTEGER	Конец связи
connection_type	INTEGER	Тип связи

Все поля в этой таблице определены целыми числовыми данными, если структура имеет одинарную связь – это единица, если двойная связь – это двойка, если присутствует тройная связь между атомами, то поле определено тройкой. Поля *id_node1* и *id_node2* указывают на наличие начальной и конечной точек связывания.

В результате модуль BioMolDiagrams позволяет автоматизировать поток информации, хранящийся в БД, и предоставить пользователю выбор показателей, которые могут быть присвоены молекулярным диаграммам, что позволяет проводить сравнительный анализ между отдельными фрагментами в структуре биополимеров.

ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Входными данными модуля BioMolDiagrams являются:

- 1) квантово-химический метод расчета, параметры оптимизации (норма градиента);
- 2) название исследуемой молекулы и принадлежность к конкретному типу фрагмента клеточной мембраны;
- 3) исходные геометрические параметры молекулы в виде z-матрицы для конкретного метода расчета;
- 4) исходные энергетические и зарядовые характеристики молекулы.

Выходными данными модуля BioMolDiagrams являются:

- 1) минимизированные энергетические параметры молекулы (общая энергия, теплота образования);
- 2) таблица оптимизированных геометрических параметров молекулы (длина химической связи);
- 3) зарядовые характеристики молекулы;
- 4) двумерная молекулярная модель (молекулярная диаграмма);
- 5) таблица входных параметров для модуля ModelInteractions в режиме квантово-химического моделирования взаимодействий и расчета основных параметров, образующихся систем.

При описании межмолекулярных взаимодействий необходимо расширить структуру базы данных. На рис. 2 приведена схема, иллюстрирующая связывание данных в таблицах, определяющих взаимодействие.

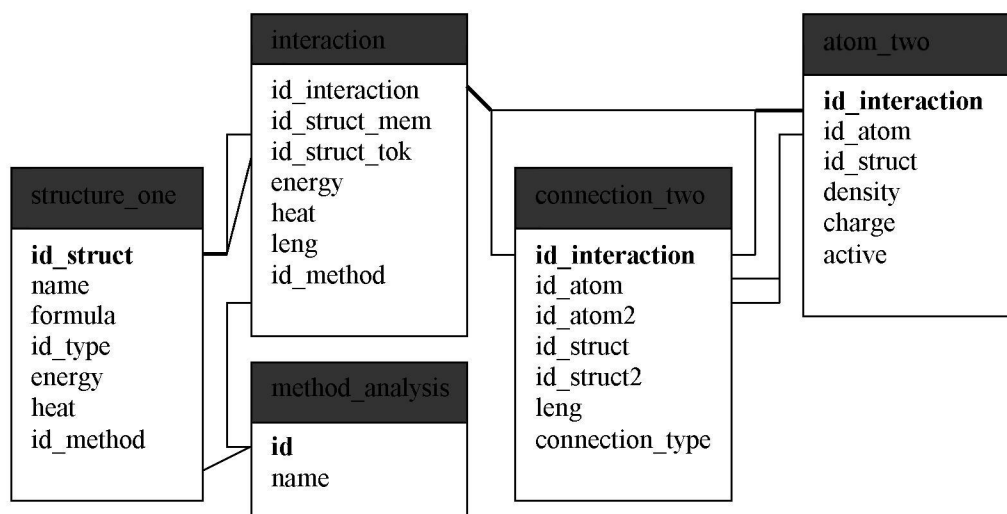


Рис. 2. Схема связей данных взаимодействующих молекул

Данные о структуре адсорбата и адсорбента приведены в табл. 10.

Таблица 10

Interaction		
id_interaction	INTEGER	Идентификатор взаимодействия
id_struct_mem	INTEGER	Идентификатор структуры – мембраны
id_struct_tok	INTEGER	Идентификатор структуры – токсиканта
energy	DOUBLE PRECISION	Энергия
heat	DOUBLE PRECISION	Количество теплоты
leng	DOUBLE PRECISION	Длина связи между структурами
id_method	INTEGER	Метод анализа

ПРИКАСПИЙСКИЙ ЖУРНАЛ: управление и высокие технологии № 2 (10) 2010

Идентификация моделей взаимодействия обеспечивается полем *id_interaction*, размер которого определен целым числом. Таблица включает идентификатор по структуре токсиканта *id_struct_tok* и по структуре компонента мембраны *id_struct_mem*. Поле *id_method* содержит информацию о методе расчета, аналогично исследованиям однокомпонентных структур. Метод анализа определяется исключительно адекватностью описания геометрии структуры, сопоставлением результатов расчетов с экспериментальными данными.

Поскольку в работе рассматриваются взаимодействия, образованные за счет водородной связи, отбор систем и внесение данных в базу производится только в том случае, если длина связи между ближайшими атомами взаимодействующих молекул лежит в интервале от 1,7–1,9 Å. Это первый критерий, по которому производится отбор адсорбционных систем. Тип данных полей *leng* числовой, а размер определен двойной, с плавающей точкой. Величина общей энергии (кДж/моль) и количество теплоты (кДж/моль) характеризуют систему в целом. Тип данных полей *energy* и *heat* определен числом, а размер соответствует двойной, с плавающей точкой.

Параметры электронной плотности и зарядовые характеристики атомов реагирующих молекул приведены в таблице 11.

Таблица 11

Atom_two

<i>id_interaction</i>	INTEGER	Идентификатор взаимодействия
<i>id_atom</i>	INTEGER	Идентификатор атома
<i>id_struct</i>	INTEGER	Идентификатор структуры
<i>density</i>	DOUBLE PRECISION	Электронная плотность
<i>charge</i>	DOUBLE PRECISION	Заряд
<i>active</i>	INTEGER	Идентификатор активности

Идентификация проводится по взаимодействию (*id_interaction*); по атому (*id_atom*); по структуре (*id_struct*), в которой находится интересующий нас атом. Тип данных, перечисленных полей, задается целым числом. Размер полей *density* и *charge* определен двойной, с плавающей точкой. Величина степени смещения электронной плотности в молекуле от одного атома к другому определена дробным значением. Атомы в системе классифицированы согласно нумерации, соответствующей одиночным структурам. Идентификатор активности (*active*) позволяет определить реакционную способность атомов при взаимодействии. Числовой тип данных, размер поля – целое число. Числовые значения зарядов и электронных плотностей атомов сгенерированы из выходного текстового файла программы Gamess, величина «ноль» в поле *active* свидетельствует о слабой активности атома; «единица» классифицирует повышенную активность атомов (графически можно представить в виде \rightarrow); «двойка» свидетельствует об активности атома к образованию не только водородной, но и химической связи (графически изображается в виде \bigcirc).

Таблица 12

Connection

<i>id_interaction</i>	INTEGER	Идентификатор взаимодействия
<i>id_atom</i>	INTEGER	Идентификатор атома
<i>id_atom2</i>	INTEGER	Идентификатор атома
<i>id_struct</i>	INTEGER	Идентификатор структуры
<i>id_struct2</i>	INTEGER	Идентификатор структуры
<i>leng</i>	DOUBLE PRECISION	Длина связи
<i>connection_type</i>	INTEGER	Тип связи

ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Характеристика длины связи в двухкомпонентных системах проводится аналогично однокомпонентным структурам, поэтому форма табл. 12 практически не изменяется.

Идентификатор взаимодействия устанавливает порядок адсорбционных систем; идентификатор атомов указывает на атомы, между которыми происходит образование связи; идентификаторы структур позволяют установить порядок взаимодействия между ними и соотнести нумерацию атомов. Тип данных характеризуется целой числовой последовательностью, а размер поля *leng* определен двойной величиной с плавающей точкой, т.е. дробным значением.

В результате модуль ModelInteractions автоматизированной системы предназначен для управления параметрами взаимодействующих систем. Это позволяет проводить отбор наиболее стабильных конфигураций, определять активные центры на поверхности биополимеров и формировать схемы взаимодействий.

Входными данными модуля ModelInteractions являются:

- 1) исходные геометрические параметры системы в виде z-матрицы;
- 2) исходные энергетические и зарядовые характеристики системы;
- 3) квантово-химический метод расчета, параметры оптимизации (норма градиента);
- 4) название и формула соединения;
- 5) длина связи между атомами взаимодействующих молекул.

Выходными данными модуля являются:

- 1) таблица оптимизированных геометрических параметров системы;
- 2) минимизированные энергетические параметры системы;
- 3) зарядовые характеристики атомов в системе;
- 4) характеристика активности атомов;
- 5) двухмерная молекулярная модель взаимодействия.

Таким образом, в рамках данного исследования была разработана структура реляционной базы данных для хранения параметров взаимодействия одно- и двухкомпонентных структур, а также интегрированная с базой данных автоматизированная система для ввода информации, расчета параметров и визуализации структур.

Библиографический список

1. *Алыков, Н. М.* Моделирование воздействия диоксида серы на структурированную белковую поверхность с использованием квантово-химического аппарата / Н. М. Алыков, Н. В. Золотарева // *Безопасность жизнедеятельности*. – 2009. – № 10. – С. 12–17.
2. *Казанцева Н. В. (Золотарева)* Теоретическое обоснование сорбции диоксида серы на структурных элементах клеточных мембран / Н. В. Казанцева (Золотарева) // *Экологические системы и приборы*. – 2007. – № 9. – С. 35–37.
3. *Казанцева, Н. В. (Золотарева)* Квантово-химическое моделирование хемосорбции диоксида серы на структурных элементах клеточных мембран / Н. В. Казанцева (Золотарева), Н. Н. Алыков // *Известия вузов*. – 2007. – Т. 50, № 12. – С. 132–133. – (Химия и хим. технология).
4. *Свидетельство* о регистрации базы данных № 2009620395. Молекулярные диаграммы структурированных поверхностей / Н. В. Золотарева, Н. М. Алыков ; заявитель и патентообладатель Астраханский государственный университет. – № 2009620306; заяв. 27.05.09; опубл. 24.07.09.
5. *Свидетельство* о регистрации программы для ЭВМ № 2009615137. Автоматизированная система Molecular Model / Н. В. Золотарева, А. Ю. Макаренко ; правообладатель Общество с ограниченной ответственностью УК «Специалист». – № 2009613882; заяв. 20.02.09; опубл. 18.09.09.
6. *Справочник биохимика* / Р. Досон, Д. Элиот, У. Элиот и др. – М. : Мир, 1991. – 544 с.