

# Современные методы молекулярного моделирования

Д.В. Зленко, П.А. Мамонов, А.М. Нестеренко

8 октября 2007 г.

## Аннотация

Методы молекулярного моделирования в настоящее время стали незаменимым инструментом изучения биомолекул, позволяющим решать широкий класс задач, зачастую недоступных экспериментальным методам. Методы молекулярного моделирования широко используются для исследования структурно-функциональных взаимосвязей в биомолекулах, поиска новых лекарств, исследования свойств отдельных молекул и их ансамблей, в том числе липидных мембран.

В рамках задачи вы ознакомитесь с теоретическими основами и вычислительными алгоритмами молекулярного моделирования. Вам будет предложено создать атомную модель молекулы аминокислоты, при помощи методов вычислительной квантовой химии рассчитать ее равновесные геометрические характеристики, а также исследовать ее поверхность потенциальной энергии. Полученные данные будут использованы для создания молекулярно-механической модели олигопептида. На заключительном этапе эта молекулярно-механическая модель будет использована для изучения ее динамических характеристик с использованием методов молекулярной динамики.

Целью работы является освоение методов квантово-химических расчетов с использованием пакета PC GAMESS и методов молекулярно-динамических расчетов с использованием пакета GROMACS. Методы квантовой химии позволяют вычислять равновесные конфигурации молекул и энергии произвольных конформаций, а также получать строгие энергетические характеристики различных движений и электронно-спиновых состояний. Методы молекулярной динамики позволяют оценивать подвижности молекул и отдельных участков молекул, а также оценивать энергию межмолекулярного и внутримолекулярного взаимодействий.

В ходе выполнения данной задачи вы научитесь работать с вышеупомянутыми пакетами программ. Задача построена таким образом, что бы показать не только принципы работы соответствующих пакетов программ, но и обучить вас применению этих принципов к исследованию произвольных молекулярных систем. В частности, таких, которые ранее не изучались численными методами.

# Оглавление

<b>1</b>	<b>Методы молекулярного моделирования</b>	<b>2</b>
1.1	Введение . . . . .	2
1.2	Молекулярная динамика . . . . .	2
1.2.1	Основы метода молдинамики . . . . .	2
1.2.2	Структура пакета GROMACS . . . . .	3
1.3	Квантовая химия . . . . .	5
1.3.1	Общие замечания о вычислительных методах . . . . .	5
1.3.2	Программа PC-GAMESS . . . . .	6
<b>2</b>	<b>Практическая часть</b>	<b>10</b>
2.1	Постановка задачи . . . . .	10
2.2	Квантово-химические расчеты . . . . .	10
2.2.1	Создание полноатомной модели аминокислоты . . . . .	10
2.2.2	Оптимизация геометрии нашей системы. . . . .	11
2.2.3	Получение необходимых характеристик системы. . . . .	12
2.3	Молекулярно-динамические расчеты . . . . .	13
2.3.1	Составление файла силового поля для валентных взаимодействий . . . . .	13
2.3.2	Составление молекулярной топологии . . . . .	14
2.3.3	Составление файла топологии . . . . .	17
2.3.4	Составление файла параметров . . . . .	17
2.3.5	Оптимизация констант жесткости . . . . .	21
2.3.6	Создание исследуемой системы и запуск молекулярной динамики . . . . .	23
2.4	Корректировка констант жесткости. Двугранные углы. . . . .	26
2.4.1	Общие замечания о топологии и корректировке . . . . .	26
2.4.2	Снятие профилей двугранных углов в квантовой химии . . . . .	26
2.4.3	Снятие профилей в молекулярной механике. . . . .	30
<b>3</b>	<b>Приложения</b>	<b>32</b>
3.1	Требования к оформлению отчета . . . . .	32
3.2	Синтаксис GROMACS . . . . .	32
3.3	Основные команды ОС Linux . . . . .	36
3.4	Основные команды текстового редактора VIM . . . . .	37

# Глава 1

## Методы молекулярного моделирования

### 1.1 Введение

В современном мире методы вычислительной квантовой химии и молекулярной динамики приобретают всё большее и большее значение. Растёт производительность вычислительных машин, увеличивается качество алгоритмов вычислений; это ведёт к росту круга задач, доступных исследованию. Растёт и доля практических задач, решаемых при помощи вычислительных методов. В рамках данной задачи большого практикума мы рассмотрим процесс расчета геометрии, силовых констант и характера молекулярных движений молекулы аланина в воде.

### 1.2 Молекулярная динамика

#### 1.2.1 Основы метода молекулярной динамики

Для расчетов молекулярной динамики применяются различные программные пакеты, как то: GROMACS, Puma, CharmM, Amber. В нашем случае мы будем говорить о пакете GROMACS. Мы будем говорить о методах молекулярной динамики на примере пакета GROMACS, однако все нижесказанное справедливо для любых реализаций МД. Этот пакет представляет собой набор программ, предназначенных для расчета траекторий движения отдельных частей молекулы, аппроксимированной механической системой физических материальных точек, связанных набором сил <sup>1</sup>

Алгоритм вычислений сводится к итерационному процессу, на каждом шаге которого:

1. По имеющемуся набору координат материальных точек (ядер атомов) вычисляются значения сил, например, кулоновские, валентные и все прочие.
2. Далее вычисляются значения ускорений для каждого центра, в соответствии со значением действующей результирующей силы и массы этого центра.
3. Затем решаются уравнения движения в предположении, что силы и ускорения не меняются во времени. Координаты центров системы для некоего, как правило, очень незначительного, времени  $\tau$  записываются в память. Это время называется *шагом интегрирования*. Новые координаты и скорости передаются на следующий шаг.

Такая механистическая модель далека от идеала, но она достаточно адекватно описывает движения молекулярных структур, если достаточно точно вычислены силовые константы и шаг интегрирования не слишком велик.

Молекулярно-динамические расчеты являются весьма ресурсоёмкими, в связи с чем размер исследуемых систем невелик. Поэтому особенно важной становится проблема построения такой границы системы, которая не вносила бы в её поведение заметных искажений. На настоящий

---

<sup>1</sup>Более подробную информацию о пакете можно получить по адресу <http://www.gromacs.org/external/online-reference-manual.html>.

момент наиболее адекватной представляется граница в виде т.н. *периодических граничных условий*. Исследуемая система окружается виртуальным параллелепипедом (“ящиком”), границы которого любая частица систему может свободно пересекать. Но при пересечении грани, частица переносится на другую сторону системы (из координаты частицы вычитается поперечный размер ящика). Т.о. система, с одной стороны, ограничена, с другой стороны граница представлена не “стенкой”, а “продолжением” исследуемой системы — её копией.

Необходимо ввести ещё одно понятие. Для ускорения расчетов, пренебрегают электростатическими и Ван-дер-Ваалсовыми взаимодействиями атомов, находящихся на большом расстоянии. Расстояние, начиная с которого взаимодействия считаются малыми и не учитываются называют “*радиусом обрезания*”. Конкретная величина радиуса обрезания определяется пользователем.

Макроскопические характеристики такой небольшой молекулярной системы сильно флуктуируют. Для поддержания среднего давления и температуры разработаны специальные алгоритмы, называемые *баростатом* и *термостатом* соответственно. Термостат вносит в систему тепловой шум, приближая распределение скоростей к нормальному Максвелловскому распределению для заданной температуры. Баростат изменяет объем ящика для создания давления, равного заданному.

В качестве частиц системы могут быть указаны как атомы так и другие частицы — в терминах разработчиков пакета — *Dummy Atoms*, виртуальные объекты соответствующие по свойствам и геометрии группам атомов, например группе —CH<sub>3</sub> или —CH<sub>2</sub>—. Введение таких частиц позволяет сократить объём вычислений.

Рассмотрим более подробно структуру пакета GROMACS.

## 1.2.2 Структура пакета GROMACS

### Типы файлов:

#### 1. Структурные.

- \*.pdb формат базы данных Protein Data Bank. Хранит координаты атомов, как полученные в ходе расчета, так и экспериментальные, например, при помощи рентгеноструктурного анализа. В последнем случае нужно иметь в виду, что незначительные ошибки в определении положения некоторых атомов могут привести к возникновению больших сил и неустойчивости модели. Координаты в этом формате записываются в ангстремах с точностью до 3-го знака.
- \*.gro внутренний формат координатного файла пакета GROMACS, хранит координаты и скорости частиц, а так же информацию о размерах расчетной области пространства — ящика, внутри «стенок» которого существует исследуемая система. Координаты записываются в нм с точностью до 4-го знака.
- \*.g96 то же, что и GRO, однако его точность вдвое выше: до 8-го знака после запятой.

#### 2. Файлы топологии.

- \*.top TOPology file. формат файлов, содержащих информацию о топологии молекулы, т.е. о всех длинах и жёсткостях всех связей, о равновесных значениях углов и их жёсткостях и т.д. Так же в \*.top файле содержится название системы, состав её (сколько и каких молекул входит в систему) и ссылки на другие файлы, содержащие, например, параметры некоторых молекул.
- \*.itp Include ToPology file, файлы содержащие геометрические и силовые параметры отдельной молекулы. Как правило, это топологии молекул, не вошедших в стандартный набор GROMACS. Расширение \*.itp имеют так же файлы силового поля, содержащие информацию, о параметрах Ван-дер-Ваальсовых взаимодействий атомов (\*nb.itp), а так же имена и параметры типов связей и углов (\*bon.itp). Пакет организован таким образом, что пользователь может создать \*.top файл содержащий в себе всю информацию о данной конкретной системе в явном виде. Или же, в случае системы состоящей из

многих различных молекул, создать \*.top файл, содержащий лишь информацию о количестве и типах молекул и ссылки на соответствующие \*.itp файлы с параметрами для молекул каждого типа. Последний вариант удобнее.

- \*.atp Atom ToPology file. файлы, хранящие информацию о типах частиц, массы и Ван-дер-Ваальсовы радиусы.
- \*.rtp Residue ToPology file. файлы со стандартными топологиями остатков (Residue Topology). Сюда входят стандартные топологии, например, аминокислот, нуклеотидов, некоторых растворителей и т.д..
- \*.hdb Hydrogen Data Base, файлы содержащие информацию о количестве и положении протонов в молекуле. Эти файлы необходимы для правильного расположения протонов, т.к. рентгеноструктурный анализ не даёт информации о их положении.

Подробнее структура файлов топологии описана ниже.

### 3. Файлы параметров или Input-файлы.

- \*.mdp Molecular Dynamics Parameters — текстовые файлы содержащие параметры расчетов, в частности, шаг интегрирования, количество шагов, параметры термо- и баростатирования и все прочие параметры модели.
- \*.tpr бинарный стартовый файл, начальные условия расчета. Является результатом объединения структурного файла, файла топологии и файла параметров.

### 4. Траекторные файлы и файлы энергий.

- \*.trr, \*.xtc бинарные файлы, содержащие координаты и скорости всех частиц системы во все моменты времени (полный и сжатый варианты, соответственно).
- \*.edr бинарный файл содержащий информацию о силах, о компонентах энергии системы и об их изменении во времени. Также в этих файлах хранится информация о температуре, давлении и поверхностном натяжении в системе.

### 5. Индекс-файлы.

- \*.ndx предназначены для передачи программам пакета GROMACS информации о группах атомов, связей или других структур, с которыми необходимо провести те или иные действия. Например, требуется выбрать из траекторного файла координаты центра масс определённой молекулы воды или все длины C—H связей. \*.ndx файлы текстовые и могут быть созданы и отредактированы вручную.

### 6. Файлы результатов.

- \*.xvg текстовые файлы, создаваемые многочисленными утилитами пакета GROMACS и содержащие различную информацию, например, автокорреляционную функцию, вычисленную для какого-либо параметра или временной ход температуры в системе.
- \*.mtx файлы матриц, хранящиеся в бинарном формате, используются, например при хранении матрицы Гессе в ходе расчета нормальных колебаний системы.

## Программы пакета GROMACS.

- \* **Подготовка структуры. Набор программ, осуществляющих преварительную обработку структурных файлов.**

**pdb2gmx:** Мощный инструмент, обрабатывающий \*.pdb файл с интересующей структурой и генерирующий для неё \*.gro и \*.top файлы. Утилита также осуществляет протонирование системы согласно имеющемуся \*.hdb файлу и создаёт ящик. При работе pdb2gmx обращается к базе стандартных топологий \*.rtp и, в случае отсутствия в ней молекул, присутствующих в системе, выдаёт сообщение об ошибке.

**editconf:** Осуществляет конвертацию структурных файлов, может создать ящик и разместить в нём исследуемую систему. Также способна повернуть, сместить и центрировать систему относительно ящика.

**genbox:** Заполняет свободное место в ящике указанным растворителем.

**\* Вычислительные программы. Программы для непосредственной установки расчета, продолжения расчета и пр..**

**mdrun:** основная программа пакета. Осуществляет расчёт молекулярной динамики системы по описанному выше алгоритму. Получает на вход бинарный \*.tpr файл, на выходе даёт траекторный файл (\*.trr) и файл энергий (\*.edr). Также на выходе записывается \*.g96 файл с финальными координатами и \*.log файл, в котором в текстовом формате записываются промежуточные результаты и другая дополнительная информация.

**grompp:** препроцессор, обрабатывает структурные файлы (\*.gro), файлы топологии (\*.top) и файл параметров (\*.mdp). На выходе даёт input-файл для mdrun (\*.tpr). Препроцессор осуществляет сборку полного файла топологии, а также осуществляет предстартовую проверку параметров модели.

**\* Обработка траекторий.**

Этот раздел составляет основную массу программ. Среди них есть программы умеющие извлекать информацию из бинарных траекторных файлов. В частности, g\_traj экстрагирует из траекторного файла координаты, скорости и силы, а g\_energy экстрагирует необходимую информацию из файла энергий. make\_ndx предназначена для создания индекс файлов.

## 1.3 Квантовая химия

### 1.3.1 Общие замечания о вычислительных методах

Квантово-химические расчеты основаны на численном решении уравнения Шредингера различными методами. Исходно первым быстрым численным методом был метод Хартри-Фока, где волновая функция системы электронов искалась как линейная комбинация атомных волновых функций. Такое приближение носит название Метода Молекулярных Орбиталей как Линейной Комбинации Атомных Орбиталей (ММО-ЛКАО). Особенность метода ХФ заключается в том, что межэлектронные взаимодействия заменяются взаимодействием каждого электрона с усредненным полем, создаваемым всеми электронами.

Из названия метода видно, что молекулярная орбиталь аппроксимируется линейной комбинацией атомных орбиталей. Базис атомных орбиталей задаются исходно пользователем и называется *базисом расчета*. Как правило, для аппроксимации атомных орбиталей выбирается в качестве базисной функции линейная комбинация гауссиан. Точность расчета в большой мере определяется величиной базиса.

Ограниченный метод Хартри-Фока пригоден для систем с замкнутыми электронными оболочками, то есть на каждой орбитали находится по два спаренных электрона. Для разомкнутых систем применяется неограниченный метод Хартри-Фока, который затрачивает вдвое больше процессорного времени, чем ограниченный.

Метод Хартри-Фока имеет два очень существенных недостатка. Во-первых, отсутствует учет энергии корреляции, или энергии Ван-дер-Ваальсовских взаимодействий. За счет усреднения поля всех электронов, корреляция движения электронов, которая может присутствовать в системе усредняется и не учитывается. Во-вторых, в таком расчете не учитываются взаимодействия возбужденных электронов с невозбужденными, вычисленные энергии возбужденных состояний не верны. Из-за этого метод Хартри-Фока нельзя использовать для расчета спектров.

Улучшить расчет в рамках Хартри-Фока можно, воспользовавшись теорией возмущений Мёллера-Плессета, которая позволяет учитывать дополнительные межэлектронные взаимодействия, не вошедшие в уравнения ХФ, как возмущение.

Существует еще два мощных метода расчета: расчет при помощи Теории Функционала Плотности (DFT) и расчет методом Много-Конфигурационного Взаимодействие (MCSCF).

Теория функционала плотности позволяет вычислить энергию системы для заданного эмпирического функционала. Метод очень удобен в применении, то есть не нужно задавать никаких специальных параметров расчета, кроме собственно функционала плотности. Расчеты с DFT позволяют корректно учесть энергию корреляции.

MCSCF, напротив, расчет очень сложный как для его корректной настройки, так и для процессора: MCSCF — самый долгий вид расчета. К нему уместно прибегать если нужна очень высокая точность вычислений и если вам необходимо рассчитать спектр молекулы. Метод основан на представлении конечной волновой функции линейной комбинацией основного и возбужденных состояний. Возможно комбинирование MCSCF и метода теории возмущений.

Иногда, наоборот требуется упрощения расчета. Существует ряд полуэмпирических методов для расчета по типу Хартри-Фока: PM3, AM1. Такой расчет занимает существенно меньше процессорного времени и хорошо подходит для оценочных расчетов.

Существуют дополнительные поправки, улучшающие расчет. К таким поправкам относится учет окружения (растворителя). Он может производиться в рамках Модели Поляризуемого Континуума (PCM) или в модели СамоСогласованного Реакционного Поля Кирквуда-Онзагера (SCRF). Оба метода конструируют полость, куда помещается система и вычисляют наведенный на непрерывном диэлектрике исследуемой системой. SCRF строит сферическую область, а PCM строит область со сложной поверхностью, образующуюся при совмещении сфер с центрами на каждом атоме системы. PCM на сегодняшний день представляется наилучшим способом учета окружения, однако если ваша система имеет большое число атомов, то необходимо самостоятельно задавать полость, что может оказаться неудобным. Кроме того в PC GAMESS PCM не поддерживается некоторыми видами расчетов, например UHF, ROHF.

Существует множество компьютерных программ, осуществляющих квантово-химические расчеты. Самые известные: Gaussian, US-GAMESS, UK-GAMESS, PC-GAMESS. Программы Gaussian и UK-GAMESS платные, мы пользуемся программами US-GAMESS и PC-GAMESS. Программа PC-GAMESS используется на наших компьютерах в непараллельном режиме, расчеты на кластере НИВЦ могут использовать до 32 процессоров одновременно.

### 1.3.2 Программа PC-GAMESS

Запуск программы pcgameSS из любого окна терминала осуществляется командой:

```
nice -19 pcgameSS -i input_file.inp -o input_file.out -f &
```

Nice -19 префикс понижения приоритета — требуется, чтобы пока расчет идет, можно было спокойно работать. Символ “&” необходим, чтобы освободить расчет от консоли. Параметры расчета и координаты системы должны быть указаны в файле `input_file.inp`, результаты расчета программа будет выводить в файл `output_file.out`.

INPUT-файл GAMESS состоит из так называемых групп. Начинается группа со слова “\$имя\_группы” и заканчивается словом “\$end”. Первый символ каждой строки INPUT-файла считается за комментарий, поэтому устанавливайте его всегда в пробел для удобства. Комментарии в INPUT-файле начинаются со знака “!”. Таким образом вы можете закомментировать любую строку, добавив перед ней “!”.

Подробно опишем все необходимые нам группы и параметры каждой:

**\$CONTROL** Задаются основные параметры расчета.

**SCFTYP** Задает тип конечной волновой функции (по умолчанию RHF).

**RHF** — расчет простым ограниченным методом Хартри-Фока, дальнейшие настройки в группе \$SCF. **ROHF** — расчет ограниченным методом Хартри-Фока для систем с незамкнутыми оболочками, т.е. пространственные части спин-орбиталей со спинами +1/2 и -1/2 не различаются. **UHF** — неограниченный метод Хартри-Фока, т.е. пространственные части спин-орбиталей со спинами +1/2 и -1/2 оптимизируются незави-



симо. **MCSCF** — расчет методом мульти-конфигурационного взаимодействия. Дальнейшие настройки в группах **\$MCSCF** и **\$DET**.

**MAXIT** Максимальное количество итераций в процедуре самосогласования. По умолчанию 20, можно увеличить, если **SCF** не сходится. Одновременно рекомендуется изменить контроль сходимости в группе **\$SCF**.

**DFTTYP** Если значение параметра задано, расчеты выполняются методом теории функционала электронной плотности. Задаёт тип функционала. Мы будем всегда пользоваться **B3LYP5** — гибридный функционал.

**RUNTYP** Задаёт цель расчета.

**ENERGY** — расчет энергии системы, нахождение молекулярных орбиталей и частичных зарядов в данной геометрии. **OPTIMIZE** — оптимизация геометрии системы. **GRADIENT** — выполняет расчет по типу **ENERGY** + расчет градиентов энергии системы по координатам ядер атомов. **HESSIAN** — выполняет расчет по типу **GRADIENT** + матрица вторых производных энергии системы по координатам ядер атомов.

**EXETYP** Указывает должна ли программа выполнить расчет или только проверить входной файл. **RUN** выполнить расчет (по умолчанию).

**CHECK** — проверить синтаксис и пр. входного файла.

**COORD** Задаёт тип координат используемых для задания положений атомов системы.

**UNIQUE** — использовать декартовы координаты, задаются только уникальные с точки зрения симметрии атомы. **CART** — использовать декартовы координаты. **ZMT** — использовать координаты **Z**-матрицы.

**ICHARG** Задаёт полный заряд системы (в зарядах электрона).

**MULT** Задаёт мультиплетность системы в соответствии с формулой  $M=2S+1$ , где **S** — спин системы.

**NZVAR** Задаёт количество степеней свободы системы за вычетом смещения и вращения системы как целого. В случае **N** частиц и нелинейной молекулы равно  $3N-6$ . Соответствующие внутренние координаты должны быть заданы в группе **\$ZMAT**. Именно они будут оптимизироваться при оптимизации геометрии системы. Если параметру присвоено значение 0 (по умолчанию) для оптимизации будут использованы декартовы координаты атомов. В последнем случае внутренние координаты можно не задавать, однако сходимость процедуры оптимизации геометрии при этом существенно ухудшается.

**\$SYSTEM** Задаются параметры касающиеся взаимодействия GAMESS с операционной системой.

**TIMLIM** Задаёт лимит времени исполнения программы в минутах.

**MWORDS** Задаёт количество миллионов слов (по 8 байт) оперативной памяти доступной программе. Всего памяти на компьютере 1024 Мб, доступную память можно посмотреть программой **gkrellm2** или **htop**.

**\$BASIS** Можно задать один из predetermined наборов базисных орбиталей который будет использован при расчетах.

**GBASIS** Задаёт тип используемого для расчетов набора базисных орбиталей. **N31** — использовать валентнорасщепленный набор контрактированных гауссовских функций. Каждая невалентная атомная орбиталь представлена линейной комбинацией **N** гауссовских функций с фиксированными коэффициентами. Каждая валентная атомная орбиталь описывается линейной комбинацией из двух функций, одна из которых представляет собой линейную комбинацию трех гауссовских функций с фиксированными коэффициентами а другая одну гауссовскую функцию. **PM3** — полуэмпирический базис, при указании такого базиса интегралы Хартри-Фока будут вычисляться неявно,

- но с помощью упрощенных схем. **MINI**, **MIDI**, **MAXI** — базисы группы японского ученого Хузинаги, урезанный, стандартный и расширенный наборы соответственно.
- NGAUSS** Задаёт количество контрактированных гауссовских функций используемых для описания невалентных атомных орбиталей при использовании валентнорасщепленных наборов базисных орбиталей (N31).
- NDFUNC** Задаёт количество поляризационных функций d-типа которые следует добавить в набор базисных орбиталей каждого тяжелого атома. Значение параметра не может превосходить 3.
- NPFUNC** Задаёт количество поляризационных функций p-типа которые следует добавить в набор базисных орбиталей каждого атома водорода. Значение параметра не может превосходить 3.
- DIFFSP** Флаг (параметр принимающий два возможных значения — **ИСТИНА** (.T.) либо **ЛОЖЬ** (.F.)), указывающий на необходимость добавить в набор базисных функций тяжелых атомов диффузные функции (с большой экспонентой).
- DIFFS** Флаг, указывающий на необходимость добавить в набор базисных функций водородов диффузные функции.
- \$GUESS** Можно выбрать алгоритм генерирующий начальные вектора разложения молекулярных орбиталей по базисным функциям которые затем оптимизируются процедурой самосогласования.
- GUESS** Значение этого параметра непосредственно задаёт используемый алгоритм **HUCKEL** использовать расширенный метод Хюккеля в базисе **MINI**. Если установить **MOREAD**, начальные орбитали будут читаться из группы **\$VEC**. Группу **\$VEC** с орбиталями можно скопировать из файла **PUNCH** предыдущих расчетов этой системы.
- NORB** Количество орбиталей. Актуально для **GUESS=MOREAD**.
- \$ZMAT** В группе задаются координаты в которых будет выполняться оптимизация геометрии и ряд других параметров.
- DLC** Флаг, указывающий использовать т.н. делокализованные координаты, представляющие собой линейные комбинации внутренних координат.
- AUTO** Флаг, указывающий автоматически сгенерировать **NZVAR** линейно независимых делокализованных координат.
- IZMAT** Массив, содержащий **NZVAR** линейно независимых внутренних координат: связей, углов, двугранных углов и пр..
- \$PCM** Если во входном файле присутствует группа **\$PCM** расчеты будут выполнены с учетом влияния на систему окружения. Учет влияния окружения производится в рамках модели непрерывной однороднополяризуемой среды. В простейшем случае достаточно указать название растворителя:
- SOLVNT** задаёт тип растворителя. Допустимые значения этого параметра: **WATER**, **NEPTANE**, **ETANOL**.
- \$STATPT** В группе задаются параметры контролирующие процедуру оптимизации геометрии.
- NSTEP** Задаёт максимальное количество шагов оптимизации геометрии (по умолчанию 20).
- METHOD** Задаёт алгоритм оптимизации, в числе которых метод Ньютона-Рафсона (**NR**), метод квадратичной аппроксимации(**QA**) и метод рациональной оптимизации функции (**RFO**).
- OPTTOL** Задаёт порог сходимости алгоритма оптимизации, по умолчанию 0.0001 Хартри.
- \$SCF** В группе производятся настройки процедуры самосогласования в методе Хартри-Фока.

**DIRSCF** Флаг, указывающий, записывать ли двухэлектронные интегралы на диск (".F.") или держать их в памяти (".T").

**SOSCF** Флаг, указывающий использовать метод SCF второго порядка для контроля сходимости.

**DIIS** Флаг, указывающий использовать метод DIIS для контроля сходимости.

**\$FORCE** Установки внутри этой группы актуальны для **RUNTYPE=HESSIAN**.

**PRTIFC** Флаг, заставляющий GAMESS выписывать матрицу Гессе во внутренних координатах.

**PURIFY** Флаг, заставляющий GAMESS исключать из гессиана поступательные и вращательные движения системы.

**DECOMP** Флаг, позволяющий указывать более 3N-6 внутренних координат.

**\$DATA** В группе задаются координаты атомов молекулярной системы. Группа имеет фиксированную структуру. В случае **COORD=UNIQUE** и отсутствия симметрии группа выглядит следующим образом:

```
$DATA
1: однострочный комментарий с названием задачи
2: точечная группа симметрии. Без симметрии - C1
3: ИМЯ_АТОМА ЗАРЯД_ЯДРА X Y Z ( в ангстремах )
4: и т.д. пока не будут заданы все атомы молекулярной системой
...
$END
```

## Глава 2

# Практическая часть

### 2.1 Постановка задачи

### 2.2 Квантово-химические расчеты

На данном этапе перед студентами ставится задача моделирования молекулы аминокислоты (аланин) методами вычислительной квантовой механики, с последующим использованием полученных результатов для создания молекулярно-механической модели аланина.

Задачи:

1. Создание полноатомной модели аминокислоты.
2. Оптимизация геометрии молекулы с учетом влияния растворителя и получение равновесных значений геометрических характеристик системы.
3. Определение парциальных зарядов на атомах и характера ковалентных связей в молекуле.
4. Исследование поверхности потенциальной энергии (ППЭ) системы:
  - определение жесткостей валентных связей и углов,
  - расчет колебательного спектра молекулы,
  - снятие профилей потенциальной энергии системы при вращении вокруг торсионных углов.

...

#### 2.2.1 Создание полноатомной модели аминокислоты

Для рисования молекул будем пользоваться программой RYMOЛ.

1. Нажимаем кнопку “Builder” и переходим таким образом в режим создания молекулы. Вы видите кнопки различных атомов на верхней панели. Конструирование аланина начинаем с метана, нажав на кнопку  $CH_4$ .
2. Теперь мы будем заменять отдельные атомы химическими группами. Например, выделим один из водородов метана и снова нажмем кнопку  $CH_4$  — мы получили этан. Аналогичным образом вы добавляете  $NH_3^+$  и  $COO^-$  группы. Чтобы удалить лишний водород на атоме кислорода, выделите атом кислорода и нажмите “-1”. Соответственно чтобы добавить водород к амино-группе, выделите азот и нажмите “+1”.
3. Готовую аминокислоту сохраните в файл с расширением PDB.
4. Теперь составьте такое соединение: к С-концу и N-концу аланина прикрепите через пептидную связь по метильной группе, то есть добавьте ацетат и метиламин на концы аланина.

5. Сохраните этот “трипептид” в другой файл.
6. Закройте PYMOL и откройте полученные вами PDB-файлы в текстовом редакторе GVIM. Обратите внимание на то, в каком порядке расположены атомы. Придайте атомам вашей системы “естественный”, по вашему мнению, порядок. Пользуйтесь командами “dd” — удалить строку, “yy” — копировать строку и “p” — вставить строку. Отредактируйте названия атомов для вашего удобства, помните, что PDB-файл чувствителен к количеству пробелов, редактируйте строки с координатами в INSERT-mode.  
Добавьте, если таковой нет строчку TITLE в начало файла. После слова TITLE напишите название вашей системы на английском языке. Удалите ненужные строки комментариев.
7. Закройте и сохраните ваши PDB-файлы. Теперь у вас есть два PDB-файла, которые содержат координаты первого приближения двух систем: аланина в состоянии цвиттер-иона и аланина в пептидном окружении.

## 2.2.2 Оптимизация геометрии нашей системы.

Все работы по постановке расчетов просьба производить в директории “ /qm”.

Для создания входного файла можно воспользоваться шаблоном “ /templates/opt.inp”. Для того чтобы конвертировать PDB-файл со структурой молекулы в формат группы \$DATA входного файла GAMESS следует воспользоваться утилитой babel:

```
babel -ipdb structure.pdb -ogamin structure.inp
```

Во вновь созданном файле structure.inp в группе \$DATA будут находиться координаты атомов в нужном формате. Копируем содержимое файла /templates/opt.inp в structure.inp, от файла structure.inp оставляем только группу \$DATA.

После того как входной файл создан, можно запустить расчет командой:

```
pcgammess -i structure.inp -o ala.opt.out
```

Не забывайте, что перед командой следует написать “nice -19”, а после поставить “&”, чтобы отсоединить расчет от терминала.

Поставьте таким образом 2 расчета с группой \$PCM и 2 расчета без \$PCM — всего вы должны поставить 4 расчета оптимизации.

Необходимо распределить ваши расчеты по четырем компьютерам. На всех четырех компьютерах ваша домашняя папка общая, поэтому просто создайте четыре разных каталога, поместите туда четыре INPUT-файла соответственно и запустите расчет на каждом компьютере. Чтобы запустить расчет на соседнем компьютере, не садясь за него, можете использовать команду:

```
ssh имя_компьютера
```

Таким образом вы оказываетесь на удаленном компьютере и можете поставить там расчет.

По завершении работы программы будет создан выходной файл ala.opt.out с результатами расчетов. Внимательно изучите выходной файл. Когда градиенты становятся меньше  $10^{-4}$  Хартри/бор (атомные единицы энергии и длины) процедура оптимизации геометрии завершается и в выходной файл выводится сообщение “STATIONARY POINT LOCATED”, за которым следуют оптимизированные координаты, начинающиеся со строки “COORDINATES OF ALL ATOMS ARE (ANGS)”. Эти координаты следует извлечь из выходного файла и создать на их основе файл в формате XYZ:

```
N
System Name
A X Y Z
...
```

N — число атомов в системе; A — символ атома; X, Y, Z — декартовы координаты атома. Затем его следует открыть программой визуального редактирования и сравнить со стартовой структурой. Запишите и геометрические характеристики (длины связей, валентные и торсионные углы)

оптимизированной системы — они понадобятся для создания топологии аланина для программы молекулярной динамики.

### 2.2.3 Получение необходимых характеристик системы.

Полученные на предыдущем этапе оптимизированные координаты мы используем для нового расчета. Мы рассчитаем колебательный спектр нашей молекулы, найдем жесткости связей, валентных углов а также извлечем из выходного файла другую информацию необходимую для написания топологии аланина для моделирования молекулярной динамики.

1. Для создания входного файла воспользуемся шаблоном “ /templates/hessian.inp”. Ниже приведен пример входного файла (см. Таблица 2.1).

```
$contrl scftyp=rhf dfttyp=B3LYP5 runtyp=hessian
  coord=unique icharg=0 mult=1 nzvar=3
! exetyp=check
$end
$system mwords=33 timlit=6000 $end
$scf dirscf=.t. diis=.t. $end
$basis gbasis=n31 ngauss=6 ndfunc=3 $end
$guess guess=huckel $end
$pcm solvnt=water $end
$zmat
  izmat(1)=1,1,2
          1,1,3
          2,2,1,3
$end
$force purify=.t. prtifc=.t. $end
$data
Water molecule
C1
O 8  0.0000000000 -0.0271237938  0.0000000000
H 1  0.7507788432  0.5484432411  0.0000000000
H 1 -0.7507788432  0.5484432411  0.0000000000
$end
```

Таблица 2.1: Листинг файла h2o\_hess.inp — получение матрицы Гессе системы.

Обратите внимание на тип расчета (runtyp = hessian) а также на группу \$zmat. В ней мы задаем массив координат, в которых будет выведена матрица Гессе вторых производных полной энергии молекулярной системы. Диагональные элементы этой матрицы являются жесткостями соответствующих степеней свободы ( $\frac{\partial^2 E}{\partial x_i^2}$ ). Эти координаты задаются в виде массива IZMAT(1). Элементы массива задаются через запятую. Первый элемент указывает тип внутренней координаты (1 — расстояние между атомами 2 — плоский угол, 3 — двугранный угол), за ним следуют номера атомов задающие соответствующую координату (два индекса в случае расстояния между атомами, три индекса в случае плоского угла, и четыре индекса в случае двугранного угла). Затем таким же образом задается следующая координата и т.д. Всего должно быть задано  $3N-6$  линейно независимых(!) внутренних координат. Допускается задание до  $3N$  внутренних координат, чтобы среди них наверняка оказались  $3N-6$  линейно независимых. Необходимо установить флаг DECOMP.

Во входной файл также добавилась новая группа \$FORCE. Флаг PRTIFC указывает указывает программе, что матрица Гессе должна быть выведена во внутренних координатах (по умолчанию она выводится в декартовых координатах). Закончив редактирование вход-

- ного файла запустите расчет, предварительно удалив оставшиеся с предыдущего запуска временные файлы PC GAMESS.
2. Запустите расчет в режиме `exetyp=check`.
  3. Если вы ввели линейно зависимые внутренние координаты, программа скажет вам об этом строкой “FOUND 27 LINEARLY INDEPENDENT INTERNALS. NEED 29.”. Это означает, что вам необходимо указать еще 2 внутренние координаты.
  4. Если внутренние координаты прошли проверку, запустите расчет с `exetyp=run`.
  5. Если вы некорректно оптимизировали геометрию вашей системы, программа напишет об этом строкой “THIS IS NOT A STATIONARY POINT. VIBRATIONAL ANALYSIS IS NOT VALID.”. В этом случае возможны два варианта: либо ваша оптимизация геометрии прошла не до конца, либо настройки расчета для оптимизации отличались от текущих.
  6. Для начала извлеките из выходного файла парциальные заряды на атомах рассчитанные по Малликену (“TOTAL MULLIKEN AND LOWDIN ATOMIC POPULATIONS”)они понадобятся при создании топологии, а также информацию о кратностях связей и валентностях атомов системы (“BOND ORDER AND VALENCE ANALYSIS”).
  7. Найдите в выходном файле матрицу Гесса во внутренних координатах (“HESSIAN MATRIX IN INTERNAL COORDINATES”)и выпишите ее диагональные элементы. Отметьте каким внутренним координатам они соответствуют.
  8. Найдите в выходном файле таблицу частот нормальных колебаний (“NORMAL COORDINATE ANALYSIS IN THE HARMONIC APPROXIMATION”)и выпишите их частоты и интенсивности. Не забудьте что первые шесть не являются нормальными колебаниями, т.к. включают смещение/вращение системы как целого. Постройте ИК-спектр системы.

Теперь для получения всей необходимой информации и создания топологии аланина нам осталось снять профили потенциальной энергии молекулы при вращении вокруг торсионных углов. Непосредственно в GAMESS сканирование ППЭ системы вдоль торсионных углов не реализовано и необходимо снять эти профили по точкам (делая для каждого значения торсионного угла расчет типа ENERGY). Эта процедура автоматизирована (см. ??).

## 2.3 Молекулярно-динамические расчеты

### 2.3.1 Составление файла силового поля для валентных взаимодействий

Геометрическая структура молекулы может быть описана набором длин связей — расстояний между атомами, значениями углов и торсионных углов (общим числом  $3N-6^1$ ). Если нас интересует не только мгновенное геометрическое состояние молекулы, а ещё и динамика изменения этой структуры, необходимо знать силовые характеристики всех перечисленных валентных взаимодействий. Если быть более точным, в пакете GROMACS необходимо определить профили потенциальной энергии для всех связей, углов и т.д.. Профили (аналитические функции) выбираются пользователем из стандартного набора (см. страницу 35).

В текстовом редакторе (GVim, Ultra Edit или любом другом) необходимо создать новый документ. Для редактирования топологии более уместно использовать GVIM, поскольку он умеет “раскрашивать” синтаксис этого типа файлов. Необходимо помнить, что символ “;” означает комментарий и всё, что написано в строке за этим символом, программами читаться не будет. Каждая строка данного файла содержит информацию о каком-либо типе связи, угла или двугранного угла. В Таблица 2.2 приведена часть файла силового поля для валентных взаимодействий.

Каждая смысловая строка данного файла начинается с ключевого слова `#define` (определить), затем — имя. Имя связи начинается с “`gb_`”, далее следует произвольная комбинация символов

<sup>1</sup>  $3N-6$  — количество степеней свободы, задаваемых жесткостями может и должно быть больше.

```
#define gb_50      2      0.11077      1.3586e+07
; H-C bond in 3-rd CH2 group of pentane
#define ga_50      2      109.3496      7.1535e+02
; H-C-H in N-methylamine
#define gi_1       2      0.0          167.42309
; planar groups 40
#define gd_50      3      10.44426      11.50823      7.74001      -22.88633      8.87117      -16.23302
; CH2-CH2-CH2-CH2 dihedral in hexane
```

Таблица 2.2: Листинг файла ffLip.itp — файла силового поля GROMACS.

(букв, цифр и подчёркиваний), как правило кодирующая химические элементы, к которым относится данная связь, а так же параметры расчетов, при помощи которых данная связь была определена. Например:

```
#define gb_CC_HEX
```

Здесь видно, что описывается связь между двумя углеродами, рассчитанная для гексана.

Далее через пробел следует тип связи. Это число, обозначающее функцию, при помощи которой описывается потенциальная энергия. Например, “1” — означает связь, аппроксимированную параболой, а “3” — потенциалом Морзе (см. Таблица 3.1). После типа связи, так же через пробелы, задаются параметры. В случае параболической связи — длина (в нм) и жёсткость (в  $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ ).

Процедура задания углов и двугранных углов аналогична. Имена углов начинаются с “ga\_”, двугранных углов с “gd\_”, неправильных двугранных углов с “gi\_”. Типы и необходимые параметры представлены в таблице. Более подробную информацию можно почерпнуть на сайте [www.gromacs.org](http://www.gromacs.org) или ТУТ. Файл необходимо сохранить в рабочую директорию с расширением itp и осмысленным названием. На данном этапе в файл силового поля необходимо записать равновесные значения для длин связей, углов и двугранных углов (“ноль” в пакете GROMACS соответствует цисоидной конформации связи). В качестве параметров жёсткости углов и связей следует использовать диагональные элементы матрицы Гессе, полученной на предыдущем этапе работы. Параметры торсионных углов прописывать не следует вовсе, но необходимо их перечислить и дать названия.

Не забывайте писать комментарии к строкам вашей топологии, давайте связям и углам осмысленные названия. Вашу топологию должно быть удобно читать.

### 2.3.2 Составление молекулярной топологии

В текстовом редакторе (GVim, UltraEdit или любом другом) необходимо создать новый документ. Ниже приведена часть файла молекулярной топологии (см. Таблица 2.3).

Теперь подробно опишем каждый пункт файла топологии:

1. Директивы препроцессора, подключающие файлы силового поля.  

```
#include "имя_файла"
```

Список возможных для GROMACS силовых полей описан ниже (см. Таблица 3.2). Здесь же вы должны подключить файл с определенными вами константами.
2. Директива [ moleculetype ]. Одна строка, содержащая имя молекулы — произвольное, желательнее до 4-ёх символов, затем в строку исключения. Исключения (nrexcl) — это число соседей данного атома, для которых НЕ будут вычисляться Кулоновские (QQ) Ван-дер-Ваальсовы (VdW) взаимодействия с данным атомом. Это необходимо в связи с тем, что атомы связаны ковалентными связями, углами и т.д., в которых уже учтены вклады VdW и QQ взаимодействий. В нашем случае nrexcl устанавливается в 3.
3. Директива [ atoms ] — атомы. Каждая строка содержит следующие параметры:
  - (a) Номер атома — по порядку от первого до n-ого.



```

#include "ffG43a2.itp"
#include "lipid.itp"
[ moleculetype ]
; Name      nrexcl
DPPC 3

[ atoms ]
;  nr      type  resnr  residu   atom      cgnr      charge   mass
   1         C      1     DPPC     C1         0       -0.3365  12.011
   2         HC      1     DPPC     H11        0        0.2250   1.008
   3         HC      1     DPPC     H12        0        0.2250   1.008

[ bonds ]
;      ai      aj
      1       2      gb_52 ; H-CH2-N
      1      13     gb_53 ; C-N

[ pairs ]
;  ai      aj  funct
   1       6    1
   1       7    1

[ angles ]
;      ai      aj      ak
      2       1       4   ga_50 ; H to CH
      2       1      13   ga_51 ; H to CN

[ dihedrals ]
;  ai      aj      ak      al
   33     36     39     42   gd_51 ; Edge dihedral, near the carboxyl
   36     39     42     45   gd_50
   28     30     79     25   gi_6

```

Таблица 2.3: Листинг файла DPPC.itp — файла топологии молекулы липида.

- (b) Тип атома. Должен соответствовать типу, обозначенному в \*.atp файле подключенного вами силового поля. Тип атома определяет его VdW взаимодействие с другими атомами.
  - (c) Номер остатка. В нашем случае — это “1”, во всех строках.
  - (d) Имя остатка — в нашем случае, одно и то же имя во всех строках (такое же как было указано выше, в директиве [ moleculetype ]).
  - (e) Имя атома. Обычно, это символ из таблицы Менделеева с добавкой — номером или значащей буквой, например, Ca — стандартное имя С-альфа атома аминокислот. А водороды, соединённые, например, с атомом C12 называют H121 и H122.
  - (f) Группа по заряду. Пользователю необходимо разбить молекулы на домены, по его мнению, обладающие некоторой обособленностью и объединить их в группы. GROMACS будет рассчитывать QQ взаимодействия между этими группами как между точечными зарядами, вне зависимости от того, на каком расстоянии находятся домены. Группы нумеруются начиная с нуля и у всех атомов, принадлежащих одной группе в соответствующей колонке должно стоять одно и то же число.
  - (g) Заряд частицы. Эффективный электрический заряд атома или dummy частицы.
  - (h) Масса. Масса атома или dummy частицы (в а.е.м.).
4. Директива [ bonds ] — связи. Каждая строка содержит следующие параметры:
- (a) Номера центров, которые соединяет данная связь.
  - (b) Имя связи. Имя связи всегда начинается с "gb\_".
5. Директива [pairs] — ближнее взаимодействие. Для атомов, между которыми взаимодействия не учитываются, благодаря прехсl, можно все же учесть взаимодействия с помощью директивы pairs. Для этого просто в столбик необходимо перечислить пары номеров атомов.
6. Директива [ angles ] — углы. Каждая строка содержит следующие параметры:
- (a) По порядку, через пробелы или табуляции — номера атомов, связанных данным углом.
  - (b) Имя угла, начинающееся с "ga\_".
7. Директива [ dihedrals ] — двугранные углы. Также здесь описываются неправильные двугранные углы. Они необходимы при описании структуры, которая не должна менять свою конформацию, например C<sub>α</sub> атом должен соответствовать L-изомеру, а не D. Каждая строка содержит следующие параметры:
- (a) По порядку, через пробелы или табуляции — номера атомов, связанных данным двугранным углом.
  - (b) Имя двугранного угла, начинающееся с "gd\_" или "gi\_".
- Описывает структуры, которые не должны менять свою конформацию, например С-альфа атом должен соответствовать L-изомеру, а не D. Каждая строка содержит следующие параметры:
- (a) По порядку, через пробелы или табуляции — номера атомов, связанных данным неправильным двугранным углом.
  - (b) Имя неправильного двугранного угла, начинающееся с "gi\_".

Файл необходимо сохранить в рабочую директорию с расширением itp и осмысленным названием.

### 2.3.3 Составление файла топологии

Файл топологии с расширением `top` — основной файл топологии, передаваемый препроцессору. В нашем случае он не будет содержать в явном виде никаких параметров, а только ссылки на них. Комментарии обозначаются точно так же, как и в `*.itp` — точкой с запятой. Ниже приведен пример основного файла топологии (см. ??).

```
#include "ffG43a2_Z.itp"
#include "DPPC_hydro.itp"
#include "SPC.itp"

[ system ]
; Name
Bilayer, 64 DPPC \& 1205 water molecules

[ molecules ]
; Compound      #mols
DPPC             64
SOL              1205
```

Таблица 2.4: Листинг файла конечной топологии — `DPPC.top`.

Таблица 2.5: `domd:dppctop`

В нашем случае `*.top` файл должен содержать следующие указания:

1. Ссылку на файл силового поля. В нашем случае `ffG43a2.itp`. Синтаксис ссылки следующий:  
`#include "ffG43a2.itp"`
2. Ссылку на файл с определёнными нами параметрами валентных взаимодействий. Синтаксис точно такой же: `#include`, и имя файла с расширением в кавычках.
3. Ссылку на файл молекулярной топологии аминокислотного остатка, созданный нами ранее. Синтаксис сохраняется.
4. Ссылку на молекулярную топологию растворителя — `"tip4p.itp"`, это топология воды, с дополнительной (четвёртой) частицей, описывающей неподеленные электронные пары кислорода. В нашем случае мы используем `"spc.itp"` — более простую модель воды. Эта строка необходима только для систем, содержащих растворитель!
5. Директиву `[ system ]`, хранящую имя исследуемой системы — абсолютно произвольная строка, без специальных знаков.
6. Директиву `[ molecules ]`, в каждой строке которой через пробелы обозначены:
  - (a) Имя молекулы, такое же, как в соответствующем файле молекулярной топологии;
  - (b) количество таких молекул в системе. Количество молекул и их названия должны точно соответствовать структурному файлу и файлам молекулярных топологий. Молекулы должны перечисляться в таком порядке, в котором они идут в `GRO/G96`-файле, это очень важно!

### 2.3.4 Составление файла параметров

В этом разделе рассматривается построение основного файла, определяющего расчеты. Файл параметров (`*.mdp`) содержит информацию о способе расчетов (алгоритме, радиусах обрезания

эффективных взаимодействий и др.), количестве и длине шага интегрирования, термостатировании, способе отображения результатов и многое другое. Мы рассмотрим только часть параметров, которые могут быть заданы в явном виде. Более подробная информация доступна на сайте [www.gromacs.org](http://www.gromacs.org) или ТУТ.

Каждая строка файла параметров начинается с ключевого слова, означающего тот или иной параметр, затем следует символ “=” и значение, характеризующее параметр (число или слово). Комментарии в файле параметров обозначаются так же точкой с запятой. Можно использовать файл параметров (Таблица 2.6) и отредактировать исходя из наших целей. Рассмотрим перечис-

```
; Common fetures
title          =          Yo
cpp            =          /bin/cpp
integrator     =          sd
ld\_seed      =          123845
dt             =          0.001
nsteps        =          10000

; Run control
nstcomm       =          10
comm_mode     =          linear
nstlist       =          10
ns_type       =          grid
rlist         =          1.0
rcoulomb      =          1.0
rvdw          =          1.0

; Output control
nstxout       =          100
nstvout       =          100
nstfout       =          100
nstlog        =          1000
nstenergy     =          100

; Temperature coupling is on in two groups
tc_grps       =          System
tau_t         =          0.1
ref_t         =          300

; Isotropic pressure coupling is now on
pcoupl        =          berendsen
pcoupltype    =          isotropic
tau_p         =          0.5
compressibility = 4.5e-5
ref_p         =          1.0
```

Таблица 2.6: Листинг файла `estim.mdp` с параметрами MD расчета

ленные в файле параметры:

### 1. Общие.

“**title**” Название системы, после знака равенства — произвольная строка.

“**cpp**” Путь до предпрепроцессора, программы, занимающейся сборкой файлов топологии и проверкой их соответствия структурному файлу (указан верный путь).

“**integrator**” Алгоритм вычислений. В нашем случае понадобятся интеграторы `sd`, `em` и `nm`, для молекулярно-динамических расчетов и анализа колебательного спектра молекул, соответственно здесь приведен пример файла предназначенного для обычных молекулярно-динамических расчетов.

“**ld\_seed**”<sup>2</sup> Стартовое число для генератора случайных чисел термостата, встроенного в интегратор. “**dt**” — шаг интегрирования в пс, оптимальны значения, от 0.5 до 0.25 фс ( $10^{-15}$  сек). “**nsteps**” — количество шагов. Должно быть таким, что бы суммарная длина траектории составила 10 нс ( $10^{-9}$  сек) или больше.

## 2. Параметры расчетов.

“**nstcomm**” Частота с которой удаляются движения центра масс, в шагах интегрирования.

“**comm\_mode**” Способ удаления движений ЦМ, в нашем случае — кодовое слово “`linear`” — т.е. удаление поступательных движений. “**nstlist**” — частота обновления списка ближайших соседей, для расчета VdW и QQ взаимодействий (в шагах интегрирования).

“**ns\_type**” Алгоритм поиска соседей, самый простой и быстрый способ (которым мы и будем пользоваться) — по сетке, кодовое слово “`grid`”.

“**rlist**” Радиус обрезания списка ближайших соседей (в нм).

“**rcoulomb**” Радиус обрезания электростатических взаимодействий.

“**rvdw**” Радиус обрезания Ван-дер-Ваальсовских взаимодействий.

## 3. Контроль выхода.

“**nstout**” Частота с которой координаты атомов записываются в файл траектории `*.trr`.

“**nstvout**” Частота с которой в файл траектории записываются скорости атомов.

“**nstfout**” Частота записи в траекторный файл сил.

“**nslog**” Частота записи промежуточных мини-отчетов в `*.log` файл.

“**nstenergy**” Частота записи энергий, температур и т.д. в энергитический файл `*.edr`. Все частоты указываются в шагах интегрирования. Важно соблюсти баланс между точностью записи координат и энергий и объемом выходных файлов: если слишком часто записывать координаты в траекторный файл, его объем может достигнуть сотен гигабайт.

## 4. Термостатирование.

“**tc\_grps**” Группы, в пределах которых следует поддерживать температуру на заданном уровне.

“**tau\_t**” Характерное время срабатывания термостата, знаменатель показателя степени члена описывающего знакпеременное трение (в пс). Чем больше число, тем более «медленным» будет термостат.

“**ref\_t**” Температура, поддерживаемая в системе (в кельвинах).

## 5. Баростатирование.

“**pcoupl**” — алгоритм баростатирования, в нашем случае по Берендсену.

“**pcoupl\_type**” — геометрия баростатирования, возможно поддержание изотропного Паскалева давления (наш случай) или нескольких видов анизотропного давления.

“**tau\_p**” — имеет тот же смысл, что и в случае термостата, это знаменатель экспоненты, описывающей поведение стенок ящика, при отклонении давления от заданного.

“**compressibility**” — сжимаемость системы, в нашем случае соответствует сжимаемости воды ( $4.5 \cdot 10^{-5} \text{Bar}^{-1}$ ). “**ref\_p**” — давление, поддерживаемое в системе (Бар). Параметр актуален для баростата Парринелло-Рахмана.

<sup>2</sup>Необходимо для полной повторяемости *in silica* эксперимента.

```
; Common fetures
title = Yo
cpp = c:\gromacs\bin\cpp
integrator = cg
nsteps = 3000

; Energy minimization stuff
emtol = 10
nstcgsteep = 100

; Run control
nstcomm = 10
ns_type = grid
rlist = 1.0
rcoulomb = 1.0
rvdw = 1.0
pcoupl = no
tcoupl = no
genvel = no
```

Таблица 2.7: Листинг файла E\_min.mdp — файла параметров для оптимизации геометрии.

Для работы понадобятся ещё два файла параметров. Один, предназначенный для оптимизации геометрии полученной системы (Таблица 2.7), другой — для проведения анализа колебательного спектра молекул (Таблица 2.8). Эти файлы отличаются от рассмотренного выше.

Приведенный файл параметров предназначен для запуска процедуры оптимизации геометрии системы. Это необходимо, во-первых, для устранения дефектов, неминуемо образующихся в ходе сборки системы. Во-вторых, оптимизация геометрии абсолютно необходима для проведения анализа нормальных колебаний, т.к. расчеты для системы, не находящейся в локальном минимуме энергии, не будут иметь физического смысла. Для оптимизации используется интегратор “cg” — Conjugate Gradient. Алгоритм вычислений сводится к последовательному изменению вектора координат атомов против направления градиента потенциальной энергии. Процесс заканчивается в локальном минимуме поверхности потенциальной энергии, что и требуется. Для его работы необходимо задать два дополнительных параметра. “emtol” — значение энергии, по достижении которого система считается оптимизированной ( $kJ \cdot mol^{-1} \cdot nm^{-1}$ ). “nstcgsteep” — частота с которой система осуществляет операцию ускоренного спуска (в шагах), это ускоряет процесс оптимизации, но несколько снижает точность. При слишком частых таких операциях возрастает вероятность локализации системы в неглубоком промежуточном минимуме. Третий и последний файл параметров, который нам понадобится, предназначен для анализа нормальных колебаний молекулярной системы. Пример такого файла приведен ниже (см. Таблица 2.8).

Интегратор “nm” не предполагает итерационных расчетов молекулярных движений. mdrun рассчитывает силовое поле для заданной конформации системы, после чего вычисляет вторые частные производные полной потенциальной энергии системы по всем координатам (вида  $\frac{\partial^2 U}{\partial x_i \partial x_j}$ ) — получается *матрица Гессе*. Диагональные элементы этой матрицы в первом приближении являются жёсткостями соответствующих виртуальных осцилляторов. Далее решается задача о связанных классических гармонических осцилляторах. Математически это можно сформулировать следующим образом: матрица Гессе и матрица кинетической энергии совместно диагонализуются путем подбора корректного координатного базиса. Компоненты такого базиса, в котором обе матрицы будут диагональными, будут нормальными колебаниями, а диагональные элементы матрицы Гессе и матрицы кинетической энергии будут жесткостями и массами независимых гармонических осцилляторов соответственно.

```

; Common fetures
title           =           Yo
cpp             =           /bin/cpp
integrator      =           nm

; Run control
ns_type         =           grid
rlist           =           1.0
rcoulomb        =           1.0
rvdw            =           1.0

```

Таблица 2.8: Листинг файла Normal\_Modes.mdp — файла MD параметров вычисления колебательного спектра системы.

### 2.3.5 Оптимизация констант жесткости

В созданном нами файле силового поля константы указаны не совсем точно. Это связано с тем, что вторые частные производные полной потенциальной энергии системы по какой-либо выделенной степени свободы не соответствуют жёсткостям, задаваемым в топологиях GROMACS. Простой пример, представьте себе метильную группу. В ней заданы 6 углов для трёх атомов водорода, три HCH и три HCC угла. Для каждого из них жёсткость есть соотв. диагональный элемент матрицы Гессе полученной в ходе КХ расчетов. Но при расчете матрицы Гессе GAMESS чуть-чуть “шевелил”, в нашем, случае, например один из HCH углов, что неминуемо влечёт за собой изменение соседних углов. Поэтому в полученной константе жёсткости содержится не только “истинная” жёсткость рассматриваемого угла, но и некоторая добавка от соседних углов и связей. Эта неточность, к сожалению, не может считаться пренебрежимо малой и от неё необходимо избавиться.

Для этого нами создана простая программа, осуществляющая подгонку параметров силового поля, так что бы частоты нормальных колебаний в GAMESS и GROMACS совпадали. Есть две версии этой программы: для MATLAB и для OCTAVE, по ряду причин предпочтительнее использовать вторую.

Для оптимизации значений констант жёсткости углов и связей необходимо запустить пакет MATLAB. В рабочей директории сохранены все необходимые файлы, вам необходимо в командной строке пакета MATLAB запустить м-файл top\_optimizer. В качестве параметров, в апострофах необходимо передать программе имена следующих файлов.

1. текстовый файл с колонкой частот нормальных колебаний, полученных в ходе КХ расчётов или измеренных экспериментально (частоты должны быть выражены в см-1).
2. файл параметров, предназначенный для оптимизации геометрии системы (см. выше).
3. стартовый структурный файл с расширением \*.gro, созданный на предыдущем этапе. 4. файл топологии, с расширением \*.top (см. выше). В данном случае он не должен содержать ссылки на файл пользовательского силового поля (строка #include "DPPC\_hydro.itp" в примере выше).
4. файл молекулярной топологии, с расширением \*.itp (см. выше), помните о том, что файл топологии (предыдущий) должен содержать ссылку именно на копию файла молекулярной топологии, созданную только что!

Для использования в линуксе был разработан более оптимизированный вариант скрипта top\_optimizer под программу GNU OCTAVE. Использование его предпочтительно, поскольку можно избежать многих “подводных камней”, которые возникают при использовании алгоритма MATLAB. Файл top\_optimizer.m содержит в самом начале множество настроек оптимизации (см. Таблица 2.9). Опишем некоторые из них:

```

%
% FREQUENCY OPTIMIZATION
%
max_step_freq = 100;           % stop after step...
step_freq     = 1e-4;         % for derivatives calculation
empiric_freq  = 010;
%
% HESSIAN OPTIMIZATION
%
max_step_hess = 000;
step_hess     = 1e-2;
empiric_hess  = 1;
%
% GLOBAL DEFINITIONS & FILENAMES
%
grad_tol      = 1 ;
em_tol        = 1e3 ;
err_tol       = 1e-5;
global gamess_file = 'ffcalc.out';
global em_mdp      = 'EnergyMinimization.mdp';
global nm_mdp      = 'NormalModes.mdp';
global start_gro   = 'metbut.gro';
global g96_file    = 'metbut.g96';
global top_file    = 'metbut.top';
global itp_file    = 'metbut.itp';
global first_freq  = '7';

```

Таблица 2.9: Верхняя настраиваемая часть файла `top_optimizer.m`, предназначенная для запуска с помощью OCTAVE.

**xxx\_freq** Оптимизация по нормальным частотам. В этом случае из OUT-файла GAMESS выбираются нормальные частоты за исключением первых **first\_freq** частот. Критерий качества топологии выбирается как схожесть частот, извлеченных из QM расчета с частотами МД-модели.

**xxx\_hess** Оптимизация по компонентам Гессиана. В этом случае ваша координаты вашей МД-системы *обязана* должны полностью совпадать с координатам QM-системы.

**max\_step\_X** Количество шагов оптимизации. Чтобы отключить оптимизацию по гессиану, просто поставьте *max\_step\_hess = 0*.

**step\_** Размер шага, влияет на точность вычисления частных производных. Нельзя выставлять очень маленьким, поскольку точность расчетов не позволяет очень точно вычислять производные и оптимизация будет расходиться.

**empiric\_** Эмпирический параметр влияющий на скорость и точность оптимизации, подбирается вручную.

**gamess\_file** OUT-файл GAMESS, расчет должен быть выполнен с `RUNTYPE=HESSIAN`.

**g96\_file** Вы должны сконвертировать свой gro- или pdb- файл в G96, при помощи программы editconf, координаты в этом формате выписываются с большей точностью.

**X\_file** Прочие файлы \*.itp, \*.top, \*.gro должны быть включены.



**first\_freq** Все частоты от первой до **first\_freq** пропускаются программой. Иногда первые одна-две частоты QM-расчета оказываются комплексными, тогда их надо исключить из сравнения. В норме мы исключаем первые 6 частот и устанавливаем значение параметра в 7, если существуют комплексные частоты, их количество надо прибавить к 7.

**X\_mdp** Параметры **em\_mdp** и **nm\_mdp** содержат ссылки на файлы параметров оптимизации геометрии и вычисления нормальных колебаний GROMACS, эти файлы вы учились создавать ранее (см. Подраздел 2.3.4).

Файл молекулярной топологии будет неоднократно перезаписан, поэтому на этом этапе работы лучше создать его копию (например **XXX\_opt.itp**). Кроме того, в этом файле необходимо в первых строках объявить варьируемые параметры. Делается это при помощи ключевого слова **#define**, точно так же, как в файле силового поля. В первых строках необходимо указать только варьируемые параметры. После списка варьируемых параметров ставится **;** **end**. Посмотрите на пример ITP-файла, правильно отредактированного для работы **top\_optimizer** (Раздел 3.2). Для MATLAB-версии программы параметры двугранных углов должны быть указаны непосредственно перед директивой **[ dihedrals ]** файла молекулярной топологии (см. выше).

После подготовки всех файлов, запустите в пакете MATLAB программу **top\_optimizer**:

```
top_optimizer( "GAMESS_freq.txt "em.mdp "start.gro "topol.top "sys_opt.itp");
```

или если вы используете OCTAVE-вариант программы:

```
octave -qf top_optimizer.m
```

После запуска вы увидите в терминале примерно следующий листинг (Таблица 2.10). Степень оптимизации можно увидеть по значению параметра **ERR**, он должен уменьшаться. Качество оптимизации проверяется величиной параметров **MAX GRADIENT** и **RMS GRADIENT**. Вы всегда можете остановить оптимизацию на определенном этапе, нажав **Ctrl+C**, однако лучше подождать завершения, поскольку программа может выбрать состояние с минимальным градиентом.

Программа **top\_optimizer** (MATLAB), по окончании расчетов выведет на экран два столбца чисел — это частоты нормальных колебаний, соответственно, полученные в **KX** расчетах и характерные для оптимизированной системы. Эти числа характеризуют аккуратность модели и их необходимо сохранить в каком-либо виде (записать в тетрадь).

После окончания процедуры оптимизации, в файл молекулярной топологии **XXX.itp**, переданном программе, будут записаны оптимизированные значения констант жёсткости. Их необходимо перенести в файл силового поля.

В нашей лаборатории были вычислены значения констант для большого количества небольших молекул, вроде глицерина, метилбутирата, метилацетата и прочих. Необходимо сравнить полученные вами константы с соответствующими константами, полученными нами ранее. Константы находятся в файле **topdb.itp** в директории **/templates**.

### 2.3.6 Создание исследуемой системы и запуск молекулярной динамики

В данном разделе описан порядок подготовки и запуска молекулярно-динамических расчетов при помощи пакета GROMACS. В качестве стартового предполагается структурный **\*.pdb** файл с исследуемой молекулой. Справка для всех программ пакета GROMACS вызывается из командной строки с использованием ключа **"-h"**, например, для препроцессора (**grompp**):

```
grompp -h
```

Результаты всегда могут быть проконтролированы визуально при помощи просмотровой программы **VMD**:

```
vmd filename.gro
```

1. Создание "ящика". Для создания рабочей области расчетов используется утилита **editconf**. Ключ **-f** — имя входного структурно файла (в нашем случае **\*.pdb**), **-o** — имя выходного

```

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
Starting TOP-OPTIMIZER utilite @
Written by Dr.Zoidberg&comcon1 @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
Processing GAMESS file...Ok
Processing ITP file...Ok
Starting optimization of topology constants.
Optimization control parameters:
=====
ERROR TOLERANCE      = 1000.000
MAXIMUM STEPS        = 100
GRADIENT TOLERANCE   = 1.000
STEP SIZE            = 1.000e-04
ERROR TYPE           = FREQ
=====

Using itp-file "metbut.itp". Optimize current set of constants:
CONSTANT #1          ga_MBT_COC 2   120.000   1.9743926447e+03
CONSTANT #2          ga_MBT_OCO 2   120.000   8.1402967364e+02
CONSTANT #3          ga_MBT_CCOc 2   120.000   9.4492787094e+02
CONSTANT #4          ga_MBT_CCOe 2   120.000   8.6683440844e+02
Starting convergance procedure:
=====
START ERROR =      1.764809e+06
STEP 1 ERR = 0.9990663477 | MAX GRADIENT = 1.244e+03 | RMS GRADIENT = 7.123e+02
STEP 2 ERR = 0.9985037982 | MAX GRADIENT = 8.127e+02 | RMS GRADIENT = 4.714e+02
STEP 3 ERR = 0.9978345875 | MAX GRADIENT = 4.862e+02 | RMS GRADIENT = 2.746e+02
STEP 4 ERR = 0.9981234894 | MAX GRADIENT = 1.253e+02 | RMS GRADIENT = 8.405e+01
STEP 5 ERR = 0.9981380705 | MAX GRADIENT = 4.183e+02 | RMS GRADIENT = 2.130e+02
STEP 6 ERR = 0.9988764124 | MAX GRADIENT = 4.572e+02 | RMS GRADIENT = 2.879e+02
[25%]

```

Таблица 2.10: Листинг работы программы top\_optimizer (OCTAVE) для метилбутирата.

структурного файла (в нашем случае \*.gro). Необходимо передать утилите размер ящика, три размера (в нм) под ключом —box, а так же передать ключ -c — означающий центрирование молекулы. Минимальный размер ящика должен быть больше чем в два раза больше, чем cutoff для QQ и VdW взаимодействий (в нашем случае больше 2 нм).

```
editconf -f filename_1 -o filename_2 -box X Y Z -c
```

2. Заливка исследуемой системы растворителем. Для дальнейших расчетов нам понадобятся как система с растворителем, так и без него, поэтому файл, полученный на предыдущем этапе, удалять не следует. Растворитель в систему добавляется автоматически, при помощи утилиты genbox. Программе необходимо передать структурный файл с исследуемой системой (ключ -cp), структурный файл с молекулами растворителя (ключ “-cs”, в нашем случае — tip4p.gro). Выходной файл — под ключом “-o”.

```
genbox -cp filename_1.gro -cs filename_2.gro -o filename_3.gro
```

3. Подготовка файлов топологии (\*.top), нам понадобятся два файла — для системы в вакууме и для системы в растворителе. Первый был создан ранее, второй отличается дополнительной строчкой со ссылкой на файл молекулярной топологии воды и строкой, содержащей информацию о количестве молекул воды в системе.

4. Создание входного файла для оптимизации геометрии. Для этого используется препроцессор `grompp`. Препроцессору необходимо передать файл параметров (`-f`), файл стартовой структуры (`-c`) и файл топологии (`-p`). На выходе (`-o`) препроцессор выдаст входной файл для основного инструмента пакета GROMACS `mdrun`. Сначала необходимо провести оптимизацию геометрии исследуемой системы. Для этого на вход программе подаётся соответствующий файл параметров и структурный файл, полученный ранее. Оптимизацию геометрии необходимо провести как для аминокислоты в вакууме, так и для аминокислоты в воде.

```
grompp -f filename_1.mdp -c filename_2.g96 -p filename_3.top -o filename_4.tpr
```

5. Запуск процедуры оптимизации геометрии. Все процедуры расчетов в пакете GROMACS осуществляются программой `mdrun`, которой на вход подаётся один единственный файл с расширением `.tpr`. На выходе `mdrun` выдаёт траекторный файл (`-o`), файл с конечными координатами атомов (структурный, `*.gro`, `-c`), файл энергий (`-e`) и лог-файл, в котором содержится отчёт с промежуточными итогами расчётов (`-g`). Сейчас нас будет интересовать выходной `*.gro` файл — хранящий координаты атомов оптимизированной системы, а так же траекторный файл (`*.trr`). Остальные файлы не имеют принципиального значения.

```
mdrun -s file_1.tpr -o file_2.trr -c file_3.g96 -e file_4.edr -g file_5.log
```

Полученные файлы нужно сохранить в рабочей директории с пометкой "`_opt`".

6. Запуск молекулярной динамики. Цель расчёта — получить карты Рамачандрана для свободной аминокислоты в растворе и вакууме, а так же наблюдение нормальных колебаний молекулы в вакууме. Отличия в запуске препроцессора от такового при оптимизации геометрии в том, что входной структурный файл — оптимизированный (получен на шаге 4). Файл параметров так же другой — предназначенный для проведения расчетов молекулярной динамики (см. п. 4). Синтаксис не отличается от шага 3. Необходимо создать два входных файла — для аминокислоты в воде и в вакууме.

Запуск `mdrun` отличается входным файлом. Необходимо помнить, что на этом этапе основными файлами, содержащими результаты расчетов, будут траекторный файл и файл энергий. Они должны получить прозрачные имена. Синтаксис запуска расчётов не отличается от шага 4. Результаты расчётов хранятся в траекторном (`*.trr`) и энергетическом файлах (`*.edr`), которые должны получить прозрачные имена.

7. Получение карт Рамачандрана. На первом этапе следует создать индекс файл, в котором были бы перечислены номера атомов, образующих интересующие нас двугранные углы. Это делается вручную в текстовом редакторе. Индекс файл в первой строке должен в квадратных скобках содержать имя группы, ниже, через знаки табуляции четвёрки номеров атомов, описывающие требуемые углы:

```
[ dihedr ]
1 2 3 4
2 3 4 6
```

Далее необходимо извлечь из файла траектории значения двугранных углов. Эту операцию осуществляет утилита `g_angle`. На вход необходимо подать: траекторный файл (`*.trr`), входной файл, с которого стартовал расчёт (`*.tpr`), индекс файл, в котором указаны требуемые двугранные углы (`*.ndx`). Так же необходимо переключить программу на работу с двугранными углами (`-type dihedral`) и попросить записать в файл все углы (`-all`). Выходной файл в формате `*xvg` (`-ov`) — текстовый файл со значениями углов. Первая колонка — время в пс, вторая — среднее значение всех углов, все последующие — конкретные значения двугранных углов. Для графического отображения карты Рамачандрана мы используем

специализированный м-файл пакета MATLAB, `gamma_plot`. Этой программе необходимо передать имя \*.xvg файла и номера колонок, содержащих нужные значения углов (сначала угла  $\phi$ , затем  $\psi$ ). В командной строке MATLAB выполните: `gamma_plot("XXX.xvg n, m)`; На выходе программа выдаёт карту Рамачандрана в плоском и объёмном виде. Необходимо построить карты Рамачандрана для аминокислоты в растворе и в вакууме и объяснить наблюдаемые различия.

## 2.4 Корректировка констант жесткости. Двугранные углы.

### 2.4.1 Общие замечания о топологии и корректировке

Чтобы корректно описать поверхность потенциальной энергии<sup>1</sup> молекулы (далее ППЭ) с помощью набора констант жесткости необходимо производить “подгонку” этих самых констант, вычисленных при помощи квантовой химии. Константы жесткости валентных связей (далее ВС) имеют высокие значения, порядка  $10^6 \text{ J} \cdot \text{nm}^{-2}$ , а диагональные элементы гесссиана, соответствующие этим константам на 1-2 порядка отличаются от недиагональных. Иными словами, взаимодействие колебания ВС с другими внутренними степенями свободы практически не наблюдается.

Иначе дело обстоит с валентными углами (далее ВУ). Диагональные элементы, соответствующие константам жесткости ВУ сравнимы по величине с недиагональными. Также необходимым вносить в топологию дополнительные углы для того, чтобы учесть общую симметрию системы. Поясним этот факт на примере молекулы этана ( $\text{C}_2\text{H}_6$ ). Молекула состоит из 8 атомов, а значит имеет  $3N-6=18$  степеней свободы: 7 валентных связей (1 C-C, 6 C-H), 1 торсионный угол (H-C-C-H) и 10 валентных углов. Очевидно, в эту десятку должны входить 6 H-C-C и 4 H-C-H. Однако у нас имеется 6 возможных H-C-H углов, и исключение любых двух приведет к тому что один или несколько водородов станут отличаться от остальных, что нарушит симметрию. Поэтому в такой топологии необходимо указывать константы для всех 6-ти H-C-H углов. Таким образом констант становится на две больше, чем  $3N-6$ .

Очевидно, константы жесткости углов в этане должны быть занижены каким-либо образом. О том, как это сделать, подробно сказано было выше (см. Подраздел 2.3.5). Совсем ничего не было сказано о торсионных углах (далее ТУ), а ведь изменения именно этих внутренних координат приводят к крупным конформационным перестройкам в белках, липидах, нуклеиновых кислотах и прочих макромолекулах. Константы ВУ и ВС рассчитывались вблизи положения равновесия, поскольку далекие от равновесия положения ввиду большой величины этих констант при комнатной температуре не достигаются. ТУ, напротив, имеют низкие эффективные константы жесткости и при комнатной температуре успевают за время расчета “пробежать” все значения.

Для ТУ, более важным является точное описание вдали от равновесного положения, то есть необходимо снимать профиль потенциальной энергии по ТУ. Это также оказывается нетрудно сделать нашими программными средствами.

### 2.4.2 Снятие профилей двугранных углов в квантовой химии

Основной вклад в энергию торсионного угла вносит электронная корреляция, по-другому — сила Ван-дер-Ваальса. Поэтому расчет энергии ТУ в программе PC GAMESS должен включать либо теорию возмущений Мёллера-Плессета, либо Теорию Функционала Электронной Плотности (DFT), либо какую-либо другую поправку, позволяющую учитывать энергию электронной корреляции (MCSCF, например). Это делается добавлением в INPUT-файл (а именно в группу \$CONTRL) одну из следующих строк:

`DFTTYPE=V3LYP5` или другой функционал. Для ТФП.

`MPLVL=2 or 4` Для теории возмущений М-П.

В PC GAMESS не реализовано сканирование энергии по ВУ и ТУ. Однако имеется программа, которая умеет производить такое сканирование, называется она ROTATOR.

ROTATOR не производит никаких расчетов, но лишь строит повернутые конформации и заставляет расчетную программу вычислять энергии, собирая значения и выписывая их в таблицы. Расчетной программой может быть GAMESS или GROMACS в зависимости от цели расчета.

Попробуем научиться пользоваться этой программой и снимем его профиль энергии ТУ НССН молекулы этана.

1. Заведите новую папку ETH\_DIN:

```
$> mkdir ETH_DIN или для пользователей mc / Krusader: F7
```

2. Восьму babel и переведи файл “opt.xyz” в файл “opt.g96”.

```
$> babel -ixyz opt.xyz -ogr96n opt_new.g96
```

3. Скопируйте из папки /template/rotator все файлы в вашу папку.

```
$>cp /template/rotator/* . Или F5 для пользователей mc / Krusader.
```

Теперь необходимо рассказать о правилах запуска этой программы. На вход этой программы подаются 3 файла: “config.ini”, “index.ndx” и ваш PDB-файл (или G96-файл). В файле index.ndx расположены настройки определения центра вращения, осей вращения и тех молекул, которые нужно вращать. Формат файла вы видите в листинге (см. Таблица 2.11).

Определение файла INDEX.NDX:

[ любое\_слово ]

atom\_n номер\_атома

(если атом является центром вращения)

номер\_атома1 номер\_атома2 номер\_атома3

(ось вращения задается ось1 = от первого до второго атома, ось2 нам не нужна)

0 0 (эти цифры нам не нужны)

номер\_молекулы

(номер(а) молекул, которые необходимо вращать)

[ end ]

Таблица 2.11: Описание определения файла INDEX.NDX для программы ROTATOR.

4. Отредактируйте ваш PDB/G96-файл, согласно примеру. Номера атомов и номера молекул указываются в “index.ndx” такими, какими они стоят в PDB/G96-файле. Вот пример PDB-файла для молекулы этана. Исходно он выглядит следующим образом:

```
TITLE          ETHANE MOLECULE
ATOM           1  C1  ETH           1      -0.752   0.001  -0.141   1.00   0.00
ATOM           2  H3  ETH           1      -1.158   0.991   0.070   1.00   0.00
ATOM           3  H4  ETH           1      -1.240  -0.737   0.496   1.00   0.00
ATOM           4  H5  ETH           1      -0.924  -0.249  -1.188   1.00   0.00
ATOM           5  C2  ETH           1       0.752  -0.001   0.141   1.00   0.00
ATOM           6  H6  ETH           1       1.158  -0.991  -0.070   1.00   0.00
ATOM           7  H7  ETH           1       0.924   0.249   1.188   1.00   0.00
ATOM           8  H8  ETH           1       1.240   0.737  -0.496   1.00   0.00
END
```

Мы будем вращать водороды (атомы 2,3,4) вокруг атома углерода (1) вдоль оси CC (5 1). Это значит, что наша молекула должны быть разбита на две молекулы. Одна, содержащая атомы 1, 2, 3, 4, 5; другая — атомы 6, 7, 8. PDB-файл должен быть изменен следующим образом:

```

TITLE      ETHANE MOLECULE
ATOM       1  C1  ETH    1    -0.752   0.001  -0.141   1.00   0.00
ATOM       2  C2  ETH    1     0.752  -0.001   0.141   1.00   0.00
ATOM       3  H3  ETH    1    -1.158   0.991   0.070   1.00   0.00
ATOM       4  H4  ETH    1    -1.240  -0.737   0.496   1.00   0.00
ATOM       5  H5  ETH    1    -0.924  -0.249  -1.188   1.00   0.00
ATOM       6  H6  ET2    2     1.158  -0.991  -0.070   1.00   0.00
ATOM       7  H7  ET2    2     0.924   0.249   1.188   1.00   0.00
ATOM       8  H8  ET2    2     1.240   0.737  -0.496   1.00   0.00
END

```

5. А index.ndx должен выглядеть так:

```

[ HCCN_ETHANE ]
atom_n 1
5 1 3
0 0
1
[ END ]

```

Согласно этому примеру, отредактируйте ваш index.ndx.

6. Теперь перейдем к файлу, определяющему остальные настройки системы, файлу “config.ini”. Ниже приведен его синтаксис (см. Таблица 2.12, красным выделены те параметры, которые мы будем изменять. Остальные параметры необходимы для корректной работы программы изменять их не рекомендуется.

ROTATOR способен работать в трех режимах: снятие энергии с помощью GAMESS, снятие энергии при помощи GROMACS и создание анимации вращающегося ТУ. Это поведение программы регулируется параметром estim\_type в группе monte\_carlo.

**estim\_type=ANIME** - создание анимации;  
**estim\_type=GAMESS** - снятие энергии GAMESS;  
**estim\_type=GROMACS** - снятие энергии GROMACS.

Вращение по ТУ задается параметрами:

**fi\_step\_x = 0.1** - шаг снятия поверхности (в радианах);  
**fi\_range\_x = [ 0, 6.28 ]** - диапазон сканирования (в радианах, за 0 принимается исходная конформация).

Каким образом названия атомов в PDB файле будут соотноситься с названиями атомов в INPUT-файле GAMESS? Эта информация указывается в файле “qmmd.ini”.

7. Вы можете создать один qmmd-файл в домашнем каталоге и добавлять в него неизвестные атомы при каждом новом расчете. Создайте файл:

```
vi ~/qmmd.ini
```

Введите в него все атомы из PDB-файла, для этого откройте окно с PDB-файлом. Напишите:

```
.:vsplit ~/ETH_DIH/origin.pdb Enter
```

Переключение между окнами Ctrl+w w. Выделите блок с названиями атомов: Ctrl+v + вправо-влево-вверх-вниз.

Скопируйте этот блок (клавиша “у”), и вставьте в ваш “qmmd.ini” (клавиша “р”) Теперь перейдите в режим редактирования (клавиша “i”) и справа от каждого имени атома напишите через пробел его обозначение в таблице Менделеева и порядковый номер вот так:

```
C1 C 6
```

```
H3 H 1
```

Выйдите из режима редактирования (ESC) и выйдите с сохранением (2 раза, поскольку открыто 2 файла).

```
:wq Enter
```

```
:wq Enter
```

Ваш qmmd.ini файл готов. Осталось прописать путь до этого файла в блоке global файла config.ini. Измените этот параметр следующим образом:

```
qm_md_map = ~/qmmd.ini
```

8. Теперь установим `estim_type = ANIME` для создания анимации и запустим программу ROTATOR:

```
$>rotator -i index.ndx -f origin.pdb -c config.ini -o tt.sf
```

9. Анимация будет лежать в файле ESTIM/output.pdb. Чтобы посмотреть её:

```
$>vmd ESTIM/output.pdb &
```

Если анимация оказывается верной, значит все параметры установлены правильно и можно перейти к запуску ROTATOR-а с `estim_type = GAMESS`, то есть к сканированию энергии по ТУ.

10. Отредактируйте группу `gamesh`, прописав там нужные вам параметры для вычисления энергий (те параметры, что там есть можно и оставить) и запускайте ROTATOR<sup>3</sup>:

```
$>rotator -i index.ndx -f origin.pdb -c config.ini -o qm.sf
```

11. Через некоторое время профиль будет готов. Когда выполнение завершится, просмотрите файл таблицы энергий:

```
$>cat qm.sf
```

Удостоверьтесь, что значения энергий реальны.

12. Чтобы нарисовать график и фитировать его, воспользуемся программой `gnuplot`:

```
$>gnuplot
>plot 'qm.sf' with lines
```

13. График выглядит как косинус, этот двугранный угол очевидно относится к простому правильному типу (см. выше). Будем аппроксимировать его соответствующей функцией:

```
>f(x)=k*(1-cos(n*x-d))
>fit f(x) 'qm.sf' using ($0*0.1):1 via k,n,d
>plot [x=0:2*pi] 'qm.sf' using ($0*0.1):1, f(x)
```

`$0*0.1` означает умножить номер строки на 0.1 и взять в качестве оси X.

14. Константы фиттирования хранятся в файле `fit.log`. Выведем последние 50 строк в файл `ethane.dih`.

```
$>tail -n 50 fit.log > ethane.dih
```

<sup>3</sup>Желательно запускать ROTATOR в отдельном окне терминала, нажав, например `Ctrl+Alt+F1`.

Отредактируйте этот файл, вычленив из него константы  $k_p$  и  $d$  и запишите в формате, пригодном для топологии (см. Подраздел 2.3.1).

### 2.4.3 Снятие профилей в молекулярной механике.

Эта процедура существенно сложнее предыдущей, зато расчет занимает десятки секунд. Перечислим по шагам основные этапы работы:

1. Создадим макет топологии нашей системы. Не будем включать туда жесткости валентных углов и связей, а также торсионные углы. В топологии должны стоять те заряды атомов, которые вы будете потом использовать, и которые были получены на предыдущем этапе работы (см. Подраздел 2.2.3).

Откроем файл топологии: `gvim ethane.itp` Как устроен файл топологии было написано выше (см. Подраздел 2.3.2). Здесь не будет групп `[angles]` и `[dihedrals]`, а будут только группа `[bonds]` и группа `[pairs]`<sup>4</sup>.

Причем в группе `[bonds]` будут проставлены только связи типа “connection”. То есть пара атомов, цифра “5”, еще пара атомов, “5” и т.д..

2. Отредактируем файл “config.ini”. Во-первых, необходимо указать `estim_type = GROMACS`. Во-вторых, нужно выбрать, какую именно энергию необходимо получать молекулярной динамикой. Если группа `[pairs]` не указана, то необходимо устанавливать `num_energy = 3`, в противном случае это число 5.
3. Запустим ROTATOR, записав поверхность в файл “md.sf”.
4. Повторите все действия с программой `gnuplot` для построения графика.
5. Теперь вычтем график, полученный молекулярной механикой, из графика, полученного квантовой химией, для этого скопируйте колонку значений из файла “md.sf” и поставьте ее справа от колонки значений “qm.sf”.
6. Теперь снова запустите `gnuplot`. И выполните следующие действия

```
>plot [x=0:2*pi] 'qm.sf' using ($0*0.1):1 title 'QM',
      'qm.sf' using ($0*0.1):2 title 'MM',
      'qm.sf' using ($0*0.1):(1-2) title 'QM-MM'
```

7. Вы нарисовали разностный профиль энергии двугранного угла. Именно его необходимо фиттировать и использовать в дальнейшем построении топологии.

```
>fit f(x) 'qm.sf' using ($0*0.1):(1-2)
```

8. Сохраним все три графика в файл в формате PNG (для вставления в отчет).

```
>set terminal png
>set output 'ethane_hcch.png'
>plot [x=0:2*pi] 'qm.sf' using ($0*0.1):1 title 'QM',
      'qm.sf' using ($0*0.1):2 title 'MM',
      'qm.sf' using ($0*0.1):(1-2) title 'QM-MM'
или вместо последней команды просто
>replot
```

<sup>4</sup>Группа `pairs` используется в особых случаях, когда поведение двугранный угол не удается описать в рамках уравнений силового поля



```
<input.rotator>
<global>
average_same      = 0
arrange_axis      = 0
fi_step_x         = 0.1
fi_range_x        = [ 0, 6.28 ]
fi_step_y         = 0.2
fi_range_y        = [ 0, 0 ]
input_efile       = input
output_efile      = output
qm_md_map         = qmmd.ini
max_errors        = 1
<global.end>
<monte-carlo>
monte_carlo = 0
estim_type = GAMESS
energy_units = kJ/mol
zero_min_energy = 1
<monte-carlo.end>
<noexec>
write_single_file = 1
<noexec.end>
<gromacs>
top_file = ../prop.top
mdp_file = /home/comcon1/POPE/ET-AMINE/DIHEDR/ener.mdp
num_energy = 1
one_mol = 1
box_size = [ 0.8, 0.8, 0.8 ]
<gromacs.end>
<gamess>
$CONTRL RUNTYP=ENERGY SCFTYP=RHF DFTTYP=B3LYP5 ICHARG=0 $END
$BASIS  GBASIS=N31 NGAUSS=6 NPFUNC=1 NDFUNC=1 $END
$SYSTEM MWORDS=10 $END
$SCF DIRSCF=.T. $END
$GUESS GUESS=HUCKEL $END
$PCM SOLVNT=ACETONE $END
$DATA
SINGLE POINT ENERGY OF ROTATED ETHANE
C1
<gamess.end>
<comment>
Finding HCCH rotation DIHEDRAL constant for
the Ethane molecule in ACETONE solvent.
6311-B3LYP5-PCM
<comment.end>
<input.rotator.end>
```

Таблица 2.12: Показательный листинг файла CONFIG.INI программы ROTATOR.

## Глава 3

# Приложения

### 3.1 Требования к оформлению отчета

### 3.2 Синтаксис GROMACS

МЕТБУТ.ИТР Пример файла топологии молекулы.

```
#define          ga_MBT_COC    2    120.000    1.9531164523e+03
#define          ga_MBT_OCO    2    120.000    7.6426447499e+02
#define          ga_MBT_CCOc   2    120.000    9.5341289870e+02
#define          ga_MBT_CCOe   2    120.000    8.6461420379e+02
;end

#define          gb_MBT_CbCa    2      0.154    7.0322736302e+06
#define          gb_MBT_HaC    2      0.110    9.4345983830e+06
#define          gb_MBT_CaCc    2      0.151    5.0175937251e+06
#define          ga_MBT_CgCbCa  2    112.130    8.8435480124e+02
#define          ga_MBT_CbCaCc  2    113.570    1.0420808366e+03
#define          ga_MBT_HCbCa  2    109.200    9.1660701170e+02
#define          ga_MBT_HCaCb  2    110.990    6.3616514212e+02
#define          ga_MBT_HCaH   2    107.660    8.9515552976e+02
#define          ga_MBT_HCaCc  2    107.660    5.1121597057e+02
#define          gb_MBT_HmC    2      0.109    1.0257723227e+07
#define          gb_MBT_OeCm   2      0.147    5.2089395661e+06
#define          gb_MBT_CcOe   2      0.138    6.5273920491e+06
#define          gb_MBT_CcOc   2      0.124    1.9405161893e+07
#define          ga_MBT_HCO    2    108.200    5.8829443593e+02
#define          ga_MBT_HCmH   2    111.050    3.1726364643e+02
#define          gb_MBT_HgC    2      0.110    1.2035934025e+07
#define          gb_MBT_CgCb   2      0.154    3.7095578851e+06
#define          gb_MBT_HbC    2      0.110    1.1302530561e+07
#define          ga_MBT_HCgC   2    111.227    4.7466161208e+02
#define          ga_MBT_HCgH   2    107.681    2.6657587767e+02
#define          ga_MBT_HCbCg  2    109.401    3.5493648891e+02
#define          ga_MBT_HCbH   2    106.095    4.2441631199e+02
;end
; dihedrals
#define gd_MBT_CCCO    3  +04.04 -00.16 -05.93 +01.51 +00.25 -00.58
#define gd_MBT_CCCC    3  +01.42 +17.87 +00.35 -19.54 +03.37 -01.25
#define gd_COO_planar  2   180.00 800.0
; derived from Methyl-Acetate
; -dih-
```

```
#define gd_MBT_CCOC 3 33.04 +09.33 +40.59 -01.54 -00.84 -03.55
#define gd_MBT_COCH 1 0 0.50 3
; -bon-
; derived from propane
; -dih-
#define gd_MBT_HCCC 1 0 3.58 3
; -bon-
[ moleculetype ]
  MBT 3

[ atoms ]
  1 C 1 MBT C1 1 -0.2637 12.00000
  2 HC 1 MBT H11 1 0.0953 1.00782
  3 HC 1 MBT H12 1 0.0953 1.00782
  4 HC 1 MBT H13 1 0.0953 1.00782
  5 C 1 MBT C2 2 -0.1803 12.00000
  6 HC 1 MBT H21 2 0.1125 1.00782
  7 HC 1 MBT H22 2 0.1125 1.00782
  8 C 1 MBT C3 3 -0.2494 12.00000
  9 HC 1 MBT H31 3 0.1572 1.00782
  10 HC 1 MBT H32 3 0.1572 1.00782
  11 C 1 MBT CC 4 0.7859 12.00000
  12 O 1 MBT OC 4 -0.6244 15.99491
  13 OA 1 MBT OE 4 -0.7440 15.99491
  14 C 1 MBT CM 5 0.0669 12.00000
  15 HC 1 MBT HM1 5 0.1280 1.00782
  16 HC 1 MBT HM2 5 0.1280 1.00782
  17 HC 1 MBT HM3 5 0.1280 1.00782

[ bonds ]
  01 02 gb_MBT_HgC
  01 03 gb_MBT_HgC
  01 04 gb_MBT_HgC
  01 05 gb_MBT_CgCb
  05 06 gb_MBT_HbC
  05 07 gb_MBT_HbC
  05 08 gb_MBT_CbCa
  08 09 gb_MBT_HaC
  08 10 gb_MBT_HaC
  08 11 gb_MBT_CaCc
  11 12 gb_MBT_CcOc
  11 13 gb_MBT_CcOe
  13 14 gb_MBT_OeCm
  14 15 gb_MBT_HmC
  14 16 gb_MBT_HmC
  14 17 gb_MBT_HmC

[ pairs ]
  1 11 1
  5 12 1
  5 13 1

[ angles ]
```

```

; -Cg-
2 1 5 ga_MBT_HCgC
3 1 5 ga_MBT_HCgC
4 1 5 ga_MBT_HCgC
2 1 3 ga_MBT_HCgH
2 1 4 ga_MBT_HCgH
3 1 4 ga_MBT_HCgH
; -Cb-
7 5 1 ga_MBT_HCbCg
6 5 1 ga_MBT_HCbCg
7 5 8 ga_MBT_HCbCa
6 5 8 ga_MBT_HCbCa
7 5 6 ga_MBT_HCbH
; -Ca-
 9 8 5 ga_MBT_HCaCb
10 8 5 ga_MBT_HCaCb
 9 8 11 ga_MBT_HCaCc
10 8 11 ga_MBT_HCaCc
 9 8 10 ga_MBT_HCaH
; -
1 5 8 ga_MBT_CgCbCa
5 8 11 ga_MBT_CbCaCc
; COO
 8 11 12 ga_MBT_CCOc
12 11 13 ga_MBT_OCO
 8 11 13 ga_MBT_CCOe
11 13 14 ga_MBT_COC
; -Cm-
15 14 13 ga_MBT_HCO
16 14 13 ga_MBT_HCO
17 14 13 ga_MBT_HCO
15 14 16 ga_MBT_HCmH
15 14 17 ga_MBT_HCmH
16 14 17 ga_MBT_HCmH

[ dihedrals ]
02 01 05 08 gd_MBT_HCCC
17 14 13 11 gd_MBT_COCH
01 05 08 11 gd_MBT_CCCC
05 08 11 12 gd_MBT_CCCO
08 11 13 14 gd_MBT_CCOC
13 11 12 08 gd_COO_planar

```

*Листинг 1. Листинг файла metbut.itp, пригодного для работы с программой top\_optimizer.*

ТИПЫ ВЗАИМОДЕЙСТВИЙ В GROMACS				
name	directive	atm	fun	constants specified
bond	bonds	2	1	$b_0$ (nm), $k_b$ ( $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ )
G96 bond	bonds	2	2	$b_0$ (nm), $k_b$ ( $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-4}$ )
connection	bonds	2	5	... nothing ...
constraint	constraints	2	1	... nothing ...
LJ/Coul. 1-4	pairs	2	1	$V(a)$ , $W(a)$
angle	angles	3	1	$\theta_0$ (deg), $k_\theta$ ( $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ )
G96 angle	angles	3	2	$\theta_0$ (deg), $k_\theta$ ( $\text{kJ} \cdot \text{mol}^{-1}$ )
proper dihedral	dihedrals	4	1	$\phi_{max}$ (deg), $k_\phi$ ( $\text{kJ} \cdot \text{mol}^{-1}$ ), <i>multiplicity</i>
improper dihedral	dihedrals	4	2	$\zeta_0$ (deg), $k_\zeta$ ( $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ )
RB dihedral	dihedrals	4	3	$C_0 - C_5$ ( $\text{kJ} \cdot \text{mol}^{-1}$ )

Таблица 3.1: Основные типы взаимодействия в GROMACS

СИЛОВЫЕ ПОЛЯ GROMACS	
название	описание
СИЛОВОГО ПОЛЯ	
ffgmh	официальное силовое поле GROMACS
ffgmh2	официальное силовое поле GROMACS с водородами (только белки)
ffG43..	GROMOS96 силовое поле для биологических и органических молекул, ориентировано в основном на расчеты в вакууме
ffopls	Optimization Potential For Liquid Simulation — силовое поля для расчетов в растворителе, имеет 1000 типов атомов, содержит определения для белков, нуклеиновых кислот и липидов
ffamber	Силовое поле группы Amber, ориентировано на расчеты НК и белков

Таблица 3.2: Основные типы взаимодействия в GROMACS

### 3.3 Основные команды ОС Linux

---

#### ОСНОВНЫЕ КОМАНДЫ

---

<b>whoami</b>	ктоя?
<b>pwd</b>	показать путь до текущего каталога
<b>cd</b> /home/student	перейти в каталог <i>/home/student</i> или домашний каталог если команда вызвана без параметров.
<b>mkdir</b> dirname	создать каталог <i>dirname</i>
<b>ls</b> /home/student	вывести содержимое каталога <i>/home/student</i> или текущего каталога если команда вызвана без параметров.
<b>ls</b> -lh /home/student	вывести содержимое каталога с дополнительной информацией о файлах.
<b>cp</b> file1.xyz file2.xyz	создать копию файла <i>file1.xyz</i> с именем <i>file2.xyz</i> . Если не указан полный путь файла операция выполняется в текущем каталоге.
<b>mv</b> file1.xyz file2.xyz	переименовать/переместить файл <i>file1.xyz</i> в <i>file2.xyz</i> . Если не указаны полные пути до файлов файл будет переименован. Если второй параметр указывает на каталог, файл будет перемещен в этот каталог.
<b>rm</b> file.xyz	удалить файл <i>file.xyz</i>
<b>rm</b> -rf dirname	удалить каталог рекурсивно и не задавать лишних вопросов.
<b>clear</b>	очистить экран.
<b>htop</b>	вызвать диспетчер задач.
<b>vim</b> file.xyz	текстовый редактор (открыть файл <i>file.xyz</i> ).
<b>man</b> cmd	отобразить справку по команде <i>cmd</i>

---

#### СТАНДАРТНЫЕ ИМЕНА КАТАЛОГОВ

---

.	текущий каталог
..	каталог уровнем выше текущего
~	домашний каталог
/	корневой каталог
/mnt/flash	ваша флешка ;)

---

#### ПЕРЕНАПРАВЛЕНИЕ ПОТОКОВ

---

<b>cmd</b> > <b>file.txt</b>	перенаправить сообщения выводимые программой в файл <i>file.txt</i> . Сообщения об ошибках выводятся на экран.
<b>cmd</b> >& <b>file.txt</b>	перенаправить все сообщения выводимые программой в файл <i>file.txt</i> .
<b>cmd1</b>   <b>cmd2</b>	перенаправить сообщения выводимые программой <i>cmd1</i> программе <i>cmd2</i> .

---

### 3.4 Основные команды текстового редактора VIM

НАВИГАЦИЯ	
<b>[[</b>	В начало документа
<b>]]</b>	В конец документа
<b>&lt;Ваше_слово&gt;</b>	Поиск Ваше_слово вперед
<b>?&lt;Ваше_слово&gt;</b>	Поиск Ваше_слово назад
<b>'&lt;символ&gt;</b>	Прыжок на закладку символ
<b>m&lt;символ&gt;</b>	Установить закладку символ
РЕЖИМЫ	
<b>Insert или i</b>	Режим редактирования
<b>Insert в i-режиме</b>	Режим замены
<b>v</b>	Режим построчного выделения
<b>Ctrl+v</b>	Режим блочного выделения
РЕДАКТИРОВАНИЕ	
<b>«символ&gt;y</b>	Копировать содержимое выделения в буфер <символ>
<b>«символ&gt;p</b>	Вставить содержимое буфера <символ> начиная с позиции курсора
<b>Ctrl+R &lt;символ&gt; в i-режиме</b>	То же самое
<b>y</b>	Копировать в буфер "
<b>p</b>	Вставить из буфера "
<b>d</b>	Вырезать в буфер "
<b>d&lt;число&gt;d</b>	Вырезать <число> строк
<b>y&lt;число&gt;y</b>	Копировать <число> строк