



Московский физико-технический институт

Факультет биологической и медицинской физики

ПРОЕКТ ПО КУРСУ ИНФОРМАТИКИ (PYTHON 3)

Визуализация и работа с химическими молекулами "MolVision"

Выполнила Щербина Елизавета Николаевна

Группа 6115

4 учебный семестр

1 Введение

Химия - это одна из основных наук естествознания, изучающая внутренний состав, внутреннее строение материи, закономерности качественных изменений, разложения и превращения веществ, а также закономерности образования новых веществ в результате качественных изменений.

Именно благодаря достижениям в этой науке человечество стремительно продолжает развиваться. С помощью неё мы не перестаем бороться с самыми страшными заболеваниями, получаем возможность сохранить и обезопасить урожай от вредителей или улучшить плодородие земли. Сложно представить, насколько химия глубоко внедрена в нашу жизнь.

На протяжении последних десятилетий химия развивается наиболее стремительно, ей занимаются в самых разных местах нашей планеты.

2 Актуальность проблемы

Из-за того, что многие лаборатории занимаются развитием химии независимо друг от друга, возникают проблемы, решение которых совсем не тривиально.

Самые актуальные из них - проблемы с спецификацией (системой правил однозначного описания состава и структуры молекулы химического вещества): группы придумывают свои варианты зашифровывания структур молекул, что не может не вызывать вопросов при кооперации между деятельными лабораториями. На данный момент существует более 50 различных спецификаций, самые популярные из которых SMILES, sdf, pdb, mol.

3 Цель проекта и суть

Целью данного проекта является возможность визуализировать молекулы. Увидеть, как выглядит тот или иной объект, оценить возможное взаимодействие между атомами, их геометрию. Модель строится в 3D формате и позволяет посмотреть на неё со всех сторон.

3.1 Почему SMILES

SMILES - Simplified Molecular Input Line Entry System, с англ. — «система упрощённого представления молекул в строке ввода». Это одна из самых распространённых в мире спецификаций, генерация которой имеет интересный смысл.

В терминах теории графов SMILES представляет собой строку, полученную путём вывода символов вершин молекулярного графа в порядке, соответствующем их обходу в глубину. Первоначальная обработка графа включает в себя удаление атомов водорода и разбивку циклов таким образом, чтобы получившийся граф представлял собой остовный лес. Местам разбиения графа ставятся в соответствие числа,

показывающие наличие связи в исходной молекуле. Для указания точек ветвления молекулы используются скобки.

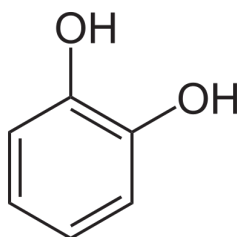
Формула SMILES может быть преобразована в двухмерную структурную формулу при помощи алгоритма (Structure Diagram Generation algorithms), разработанного Хелсоном. Преобразование не всегда даёт однозначный результат. Преобразование в трехмерную структурную формулу производится с использованием принципа минимальной энергии образования вещества.

3.2 Действие

- Программа предлагает ввести нужную молекулу в формате SMILES
- Выводит её координаты в терминал
- Рисует в новом окне 3D-структуру молекулы

4 Пример

В качестве примера нарисуем молекулу пирокатехина (SMILES: Oc1ccccc1O), имеющую следующий вид:



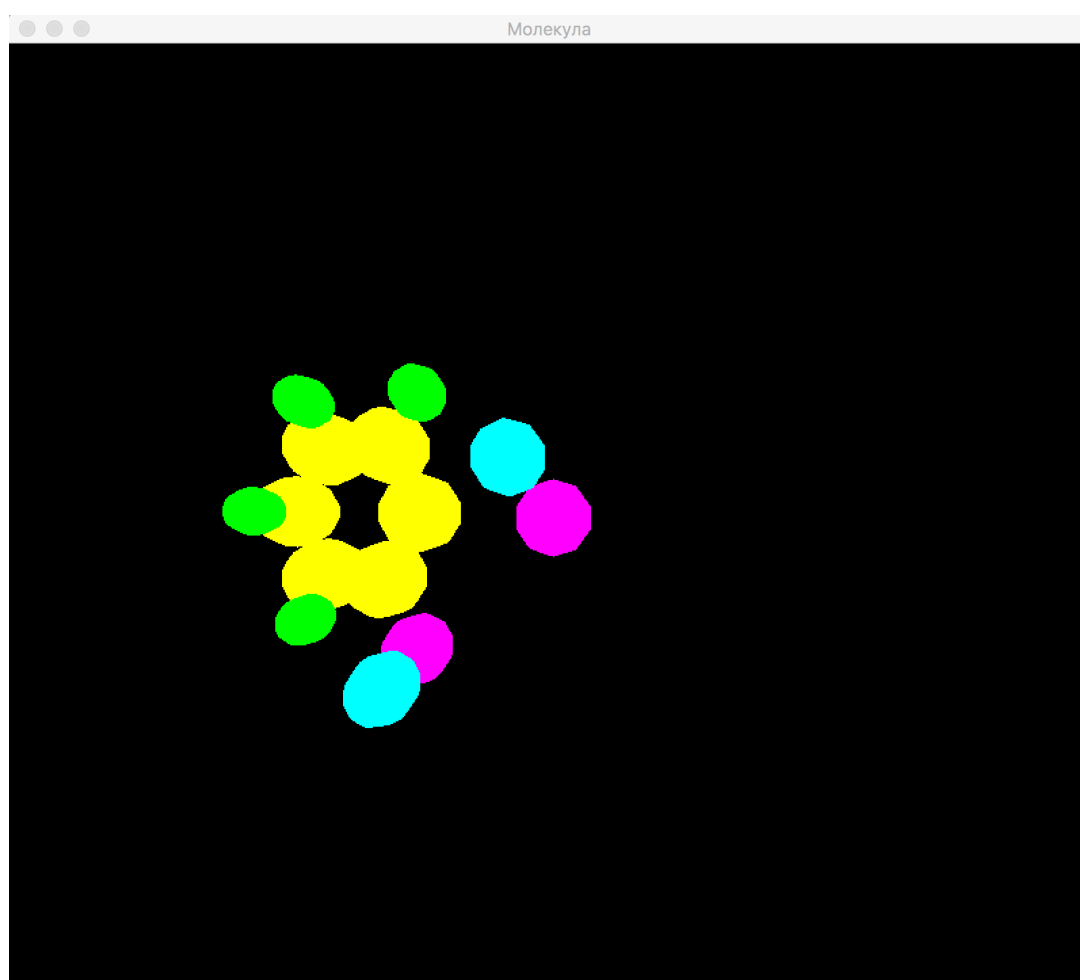
Запускаем программу:

```
Рабочий стол — python3 MolVision.py — 80×24
[net127:Desktop apple$ python3 MolVision.py
Введите молекулу в формате SMILE (например, пропилен -> C=CC): ]
```

Вводим молекулу пирокатехина и получаем нужные для построения координаты:

```
Рабочий стол — python3 MolVision.py — 80×24
[net127:Desktop apple$ python3 MolVision.py
Введите молекулу в формате SMILE (например, пропилен -> C=CC):Oc1ccccc1O
O3 2.315841784746324e+77 2.315841784746324e+77 0.06482431012864291
Car 2.3241933590152017 -0.03610934823187167 -0.02746540461731621
Car 3.0006192158530887 1.1830199025815884 -0.05895431898301041
Car 4.3868584336537335 1.1945739890604212 -0.23482049390545404
Car 5.094268579312188 -0.004939592400327703 -0.3110546600673708
Car 4.4211444841054774 -1.2253356851946162 -0.21050801263790692
Car 3.0308192376626373 -1.2339821365997385 -0.10227865035288604
O3 2.3475785359595602 -2.418893325156786 -0.08106553038844433
HO 0.6189922106349526 0.8368832658090094 -8.935921857821305e-05
H 2.470643824028208 2.122865436863821 0.04659659980059887
H 4.915748014175216 2.142087683560894 -0.3096025770363556
H 6.173446432248721 0.009083358104725941 -0.44623095099302484
H 4.989831989132115 -2.14824943300096 -0.22205669803001496
HO 2.9674853722417964 -3.1432939548812966 -0.24587301797843034
]
```

Наша полученная модель молекулы (немного размазанная, т.к. молекула находится в движении):



5 Обсуждение и вывод

В ходе данной работы мы научились работать с классами молекул, рисовать объекты с помощью известной и многофункциональной библиотеки OpenGL, познакомились с методами работы в OpenBabel (Pybel - её составляющая), вспомнили работу с двумерными массивами, особенности Python3.

Также с помощью этой программы можно загружать файл полностью, без ввода. Можно поменять формат ввода и дополнить его.

Здесь приблизительно сохранены размеры молекул, группы одинаковых молекул раскрашены одним цветом, что позволяет сделать выводы о геометрии и возможных активных частей молекул.

Данная работа лишь начало более серьёзной программы для хемоинформатики, работы с разными спецификациями, считывании и визуализации молекул. Поле для разработки ещё очень большое.

6 Используемые библиотеки

В ходе данной работы были использованы несколько дополнительных библиотек:

- Pygame
- OpenGL
- Pybel