



APLITE

Версия 0.44-0.50

Руководство пользователя

Редакция 1.3

Содержание

1. Назначение программы	4
2. Запуск программы	4
2.1. Запуск в режиме смены формата молекулы.....	4
2.2. Запуск в режиме расстановки водородов.....	5
2.3. Запуск в режиме разделения базы молекул на отдельные файлы молекул.....	5
2.4. Запуск в режиме объединения файлов молекул в базу молекул.....	6
2.5. Запуск в режиме записи координат центра лиганда в файл параметров для SOLGRID.....	6
3. Форматы входных и выходных файлов	7
4. Описание параметров программы APLITE	7
4.1. Наиболее важные параметры.....	7
Version	7
AromaticBonds	7
ProtonateType	8
4.2. Параметры расстановки водородов на белок.....	8
AlarmAldehydeGroup	8
ConvertAldehyde2Carboxyl	8
FlipHistidine	8
4.3. Параметры оптимизации положений водородов.....	8
RadiusOfInfluence	8
NRotHPositions	9
LBFSSStepsForH	9
SilentHOptimization	9
4.4. Параметры чтения PDB файлов.....	9
IgnoreVariantAtomPosition	9
IgnoreHETATM	9
DeleteWater	10
4.5. Параметры типизации и вычисления энергии.....	10
Epsilon	10
FeCharge	10
CuCharge	10
MaxRingSize	10
EBLGap	10
EACos3	11
EACos4	11
EACos	11
4.6. Внутренние параметры оптимизатора.....	11
DecreaseStep	11
IncreaseStep	11

A0Start	11
LittleA0_LBFGS	12
LBFGS_M	12
restart_area	12
n_crit_restart	12
4.7. Пример содержимого файла параметров	12
1. Назначение программы	4
2. Запуск программы	4
2.1. Запуск в режиме смены формата молекулы.....	4
2.2. Запуск в режиме расстановки водородов.....	5
2.3. Запуск в режиме разделения базы молекул на отдельные файлы молекул	5
2.4. Запуск в режиме объединения файлов молекул в базу молекул.....	6
2.5. Запуск в режиме записи координат центра лиганда в файл параметров для SOLGRID	6
3. Форматы входных и выходных файлов	7
4. Описание параметров программы APLITE	7
4.1. Наиболее важные параметры	7
4.2. Параметры расстановки водородов на белок.....	8
4.3. Параметры оптимизации положений водородов.....	8
4.4. Параметры чтения PDB файлов	9
4.5. Параметры типизации и вычисления энергии	10
4.6. Внутренние параметры оптимизатора.....	11
4.7. Пример содержимого файла параметров	12

1. Назначение программы

Программа APLITE предназначена для:

- 1) расстановки водородов на лигандах и белках;
- 2) конвертации форматов файлов молекул с типизацией и расстановкой зарядов;
- 3) разделения базы молекул на отдельные молекулы;
- 4) слияния отдельных молекул в базу молекул;
- 5) определение геометрического центра лиганда и запись его координат в формате файла параметров для SOLGRID

2. Запуск программы

2.1. Запуск в режиме смены формата молекулы

Строка запуска:

```
aplite t2xyz [parameter_file] input_molecule
        c2xyz
        t2mol
        c2mol
        t2hin
        c2hin
        t2mrk
        c2mrk
        t2pdb
        c2pdb
        t2sdf
        c2sdf
        t2mol2
        c2mol2
```

Первый аргумент – ключевое слово, определяющее, в какой формат и как следует перевести молекулу `input_molecule`. Вторым, необязательным, аргументом – это файл с параметрами `aplite`. Последний аргумент – имя файла с исходной молекулой, которая также может быть в любом из этих семи форматов.

Если ключевое слово начинается на «t», то в процессе смены формата также проводится присвоение атомам MMFF-типов и зарядов и, возможно, вносятся некоторые коррекции в структуру.

Если ключевое слово начинается на «с», то структура молекулы по возможности сохраняется неизменной. Изменения порядка или вида связей так же не произойдет (то есть опция `AromaticBonds delete` или `AromaticBonds create` не работает).

Имя выходного файла совпадает с именем входного, отличаясь только расширением, которое соответствует новому формату. Если молекула переводится в собственный же формат, то входной файл перезаписывается выходным. Это может понадобиться для расстановки MMFF-типов и зарядов.

Пример запуска 1:

```
./aplite-0.44-i64i t2mol ligand.hin
```

В данном случае на выходе получится файл `ligand.mol` с той же самой молекулой, что и файл `ligand.hin`

Пример запуска 2:

```
./aplite-0.44-i64i t2mol ligand.mol
```

Файл `ligand.mol` перезапишется новым, в котором будут указаны MMFF-типы атомов и заряды.

2.2. Запуск в режиме расстановки водородов

Строка запуска:

```
aplite h2xyz h2mol h2hin h2mrk param.txt input_molecule h2pdb h2sdf h2mol2
```

Первый аргумент – это ключевое слово, задающее формат выходного файла.

Второй аргумент – это имя файла с параметрами. В этом файле параметров указывается, в частности, по какому алгоритму расставлять водород – как на белок или как на лиганд.

Третий аргумент – это входная молекула, на которую требуется расставить водород.

Выходной файл имеет имя формата `input_molecule_h` с соответствующим расширением.

Пример запуска:

```
./aplite-0.44-i64i h2mrk param.txt protein.pdb
```

На выходе создается файл `protein_h.mrk` с расставленными водородами.

2.3. Запуск в режиме разделения базы молекул на отдельные файлы молекул

Строка запуска:

```
aplite s2xyz s2mol s2hin s2mrk [parameter_file] input_database s2pdb s2sdf s2mol2
```

Первый аргумент – это ключевое слово, определяющее формат выходных файлов молекул.

Второй, необязательный, аргумент – это файл с параметрами `aplite`.

Последний аргумент – это имя базы молекул в формате (и с соответствующим расширением) `SDF` или `MOL2`.

В процессе работы создается директория с именем, как у входной базы молекул (только без расширения) и туда записываются отдельные файлы молекул в формате, определяемым ключевым словом.

Имена этих молекул:

- 1) Берутся из базы молекул
- 2) Если в базе молекул отсутствует имя, то берется порядковый номер молекулы в базе
- 3) Если такое имя уже было, то к нему прибавляется окончание «_02», «_03» и так далее (до «_99»).

Пример запуска:

```
./aplite-0.44-i64i s2mrk aminiacids.sdf
```

В текущем каталоге будет создан подкаталог `aminiacids` и туда будут записаны молекулы из базы `aminiacids.sdf` в формате `MRK`.

2.4. Запуск в режиме объединения файлов молекул в базу молекул

Строка запуска:

```
aplite j2sdf [parameter_file] input-directory
```

Первый аргумент – ключевое слово.

Второй, необязательный, аргумент – это файл с параметрами `aplite`.

Последний аргумент – это имя каталога, файлы из которого надо объединить в базу молекул.

После работы будет создан `.sdf` файл с именем, совпадающим с именем входного каталога, содержащий объединение все файлов молекул из входного каталога. Имена молекул в базе молекул совпадают с именами файлов молекул без их расширения.

2.5. Запуск в режиме записи координат центра лиганда в файл параметров для SOLGRID

Строка запуска:

```
aplite par4grid input_ligand
```

Первый аргумент – ключевое слово.

Второй аргумент – имя файла с лигандом.

После работы программы будет создан `.rag` файл, совпадающий по имени без расширения с файлом лиганда, в котором будут записаны координаты центра лиганда в формате, необходимом для использования в `SOLGRID 2.01`.

3. Форматы входных и выходных файлов

Программа APLITE может считывать и записывать файлы молекул в форматах XYZ, MOL, HIN, MRK, PDB, SDF, MOL2. При работе с SDF и MOL2 форматами как с единичными молекулами при чтении считывается первая молекула, при записи записывается единственная молекула.

Программа APLITE может считывать базы молекул в форматах SDF и MOL2, а записывать – в формате SDF

4. Описание параметров программы APLITE

Все параметры работы программы APLITE записываются в текстовом файле в произвольном порядке. Каждый параметр описывается одной строкой. Часть строки, отделенная от начала символами «//» (без кавычек), является комментарием и не влияет на работу программы. Формат строки параметров: ключевое слово, перед которым не должно быть никаких посторонних символов, затем произвольное количество пробелов (больше нуля), затем значение (значения) этого параметра. Если какой-либо параметр не приводится в файле параметров, то ему присваивается значение по умолчанию. Нельзя писать имя параметра и не указывать значение для него. В имени параметра не может быть пробелов. Регистр символов в имени параметра существенен.

Значение параметра может быть целым числом, вещественным числом или строкой (без пробелов внутри или же заключенной в кавычки ").

4.1. Наиболее важные параметры

Version

Строковый параметр, определяющий, для какой версии программы предназначен этот файл параметров. При несовпадении с версией программы или отсутствии параметра программа завершает свою работу.

Пример использования: `Version 0.44`

Значение по умолчанию: `0.00`

Диапазон значений: строки вида `a.bc`, где `a`, `b`, `c` – цифры.

AromaticBonds

Строковый параметр, определяющий, в каком виде представлять ароматические кольца – со связями ароматического типа или с чередующимися одинарными-двойными связями.

Пример использования: `AromaticBonds delete`

Значение по умолчанию: `keep`

Диапазон значений: одно из трех слов:

`delete` (преобразовать все связи к чередующимся одинарным-двойным)

`keep` (оставить тип связей без изменения)

`create` (преобразовать все связи к ароматическому типу)

ProtonateType

Строковый параметр, определяющий режим расстановки водородов на соединение: как на белок или как на лиганд. Может принимать только два значения: PROTEIN (для режима расстановки водородов как на белок) или LIGAND (для режима расстановки водородов как на лиганд).

Режим расстановки водородов на лиганд пока не реализован.

Пример использования: `ProtonateType PROTEIN`

Значение по умолчанию: PROTEIN

Диапазон значений: PROTEIN или LIGAND

4.2. Параметры расстановки водородов на белок

AlarmAldehydeGroup

Целочисленный параметр, определяющий, надо ли выводить предупреждение при обнаружении альдегидных групп. Обычно альдегидные группы появляются при потере второго кислорода в терминальном карбоксиле аминокислотной цепи.

Пример использования: `AlarmAldehydeGroup 1`

Значение по умолчанию: 1

Диапазон значений: 0 или 1

ConvertAldehyde2Carboxyl

Целочисленный параметр, определяющий, надо ли преобразовывать альдегидные группы в карбоксильные. Обычно альдегидные группы появляются при потере второго кислорода в терминальном карбоксиле аминокислотной цепи.

Пример использования: `ConvertAldehyde2Carboxyl 1`

Значение по умолчанию: 0

Диапазон значений: 0 или 1

FlipHistidine

Целочисленный параметр, определяющий, надо ли пробовать различные варианты положения водорода в гистидине.

Пример использования: `FlipHistidine 1`

Значение по умолчанию: 1

Диапазон значений: 0 или 1

4.3. Параметры оптимизации положений водородов

RadiusOfInfluence

Вещественный параметр, определяющий размер области в пространстве вокруг каждого водорода, внутри которой прочие атомы влияют на положение этого водорода. Чем он меньше, тем быстрее проходит оптимизация.

Пример использования: `RadiusOfInfluence 7.0`

Значение по умолчанию: 9.0

Диапазон значений: от 5.0 до `+infinity`

NRotHPositions

Целочисленный параметр, определяющий, сколько пробуются положений при вращении групп с водородами. Чем он меньше, тем быстрее проходит оптимизация.

Угол одного шага поворота равен $2 \cdot \pi / \text{NRotHPositions}$.

Пример использования: `NRotHPositions 67`

Значение по умолчанию: 67

Диапазон значений: от 3 до `+infinity`

LBFGSStepsForH

Целочисленный параметр, определяющий количество шагов оптимизации положения водородов.

Пример использования: `LBFGSStepsForH 20`

Значение по умолчанию: 100

Диапазон значений: от 1 до `+infinity`, 0 значит неограниченное количество.

SilentHOptimization

Целочисленный параметр, определяющий надо (0) или нет (1) выводить лог оптимизации положений водородов.

Пример использования: `SilentHOptimization 0`

Значение по умолчанию: 0

Диапазон значений: 0 или 1

4.4. Параметры чтения PDB файлов

IgnoreVariantAtomPosition

Целочисленный параметр, определяющий, как поступать с множественными положениями атомов в PDB файле.

0 – останавливать программу, если есть множественные положения атомов,

1 – игнорировать все множественные положения атомов и использовать первое.

Пример использования: `IgnoreVariantAtomPosition 1`

Значение по умолчанию: 1

Диапазон значений: 0 или 1

IgnoreHETATM

Целочисленный параметр, определяющий надо (1) или нет (0) игнорировать записи HETATM в PDB-файле.

Пример использования: `IgnoreHETATM 0`

Значение по умолчанию: 0

Диапазон значений: 0 или 1

DeleteWater

Целочисленный параметр, определяет, надо ли игнорировать молекулы воды (остатки с сименами «НОН») при чтении PDB файла.

Пример использования: `DeleteWater 1`

Значение по умолчанию: 1

Диапазон значений: 0 или 1

4.5. Параметры типизации и вычисления энергии

Epsilon

Вещественный параметр, диэлектрическая проницаемость среды.

Пример использования: `Epsilon 1.0`

Значение по умолчанию: 1.0

Диапазон значений: от 0.0 до +infinity

FeCharge

Целочисленный параметр, заряд на ионе железа. Определяет, какой заряд будет присвоен ионам железа..

Пример использования: `FeCharge 2`

Значение по умолчанию: 2

Диапазон значений: 2 или 3

CuCharge

Целочисленный параметр, заряд на ионе меди. Определяет, какой заряд будет присвоен ионам меди.

Пример использования: `CuCharge 2`

Значение по умолчанию: 2

Диапазон значений: 1 или 2

MaxRingSize

Целочисленный параметр, максимальный размер анализируемых кольцевых структур (максимальное количество атомов в кольце). Чем он больше – тем дольше анализируется молекула.

Пример использования: `MaxRingSize 16`

Значение по умолчанию: 16

Диапазон значений: от 6 до +infinity

EBLGap

Вещественный параметр, определяющий максимальное отклонение длины связи от равновесной в ангстремах. При выходе за допустимый диапазон выводится предупреждение. Также используется при восстановлении матрицы связностей при чтении XYZ файлов.

Пример использования: `EVLGap 0.2`
Значение по умолчанию: `0.2`
Диапазон значений: от `0.0` до `+infinity`

EACos3

Вещественный параметр, определяющий максимальный косинус валентных углов в трехчленных циклах.

Пример использования: `EACos3 0.6`
Значение по умолчанию: `0.6`
Диапазон значений: от `0.0` до `1.0`

EACos4

Вещественный параметр, определяющий максимальный косинус валентных углов в четырехчленных циклах.

Пример использования: `EACos4 0.3`
Значение по умолчанию: `0.3`
Диапазон значений: от `0.0` до `1.0`

EACos

Вещественный параметр, определяющий максимальный косинус прочих валентных углов.

Пример использования: `EACos 0.05`
Значение по умолчанию: `0.05`
Диапазон значений: от `0.0` до `1.0`

4.6. Внутренние параметры оптимизатора

DecreaseStep

Вещественный параметр, множитель для уменьшения шага при одномерной оптимизации.

Пример использования: `DecreaseStep 0.2`
Значение по умолчанию: `0.2`
Диапазон значений: от `0.0` до `1.0`

IncreaseStep

Вещественный параметр, множитель для увеличения шага при одномерной оптимизации.

Пример использования: `IncreaseStep 1.6`
Значение по умолчанию: `1.6`
Диапазон значений: от `1.0` до `+infinity`

A0Start

Вещественный параметр, величина первого пробного шага в первой итерации.

Пример использования: `A0Start 1.0`
Значение по умолчанию: `1.0`
Диапазон значений: от `0.0` до `+infinity`

LittleA0_LBFGS

Вещественный параметр, наименьшая величина шага при одномерной оптимизации. Если пробный шаг станет меньше этой величины, то одномерная оптимизация считается неудавшейся.

Пример использования: `LittleA0_LBFGS 1.0e-5`
Значение по умолчанию: `1.0e-5`
Диапазон значений: от `0.0` до `+infinity`

LBFGS_M

Целочисленный параметр, количество предыдущих шагов оптимизации, информация о которых хранится в оперативной памяти.

Пример использования: `LBFGS_M 8`
Значение по умолчанию: `8`
Диапазон значений: от `1` до `+infinity`

restart_area

Целочисленный параметр, размер области для отслеживания неудавшихся шагов оптимизации.

Пример использования: `restart_area 251`
Значение по умолчанию: `251`
Диапазон значений: от `1` до `+infinity`

n_crit_restart

Целочисленный параметр, порог количества неудавшихся шагов оптимизации за последние `restart_area` шагов. При превышении этого числа оптимизация останавливается.

Пример использования: `n_crit_restart 51`
Значение по умолчанию: `51`
Диапазон значений: от `1` до `n_crit_restart`

4.7. Пример содержимого файла параметров

```
//set for protonation 3DWB_c
Version 0.44
AromaticBonds delete
ProtonateType PROTEIN
AlarmAldehydeGroup 1
ConvertAldehyde2Carboxyl 1
FlipHistidine 1
RadiusOfInfluence 13.0
NROTHPositions 67
LBFGSStepsForH 120
```

```
SilentHOptimization 0
IgnoreVariantAtomPosition 1
IgnoreHETATM 0
DeleteWater 1
Epsilon 1.0
FeCharge 2
CuCharge 2
MaxRingSize 16
EBLGap 0.2
EACos3 0.6
EACos4 0.3
EACos 0.05
DecreaseStep 0.2
IncreaseStep 1.6
A0Start 1.0
LittleA0_LBFGS 1.0e-5
LBFGS_M 8
restart_area 251
n_crit_restart 51
```