

## Компьютерные программы для генерации конформаций лигандов – потенциальных лекарственных препаратов

М.И. ШАЛАДОНОВА

В статье проведена сравнительная характеристика программ, используемых при создании конформаций лигандов. На основании проанализированных данных выбрана онлайн-программа FROG2 ввиду ее следующих характеристик: доступности, способности генерировать конформации для циклических молекул, стереоизомеров, удобства интерфейса; высокой скорости генерации конформаций. Она позволяет быстро проанализировать конформационное пространство лигандов. Результат работы программы можно сохранить в виде определенных файлов наиболее часто используемых форматов, которые могут использоваться как входные данные для других программ, позволяющих рассчитать дескрипторы химической структуры. Полученные дескрипторы необходимы для дальнейшей оценки биологической активности соединения – потенциального активно действующего вещества при разработке нового лекарственного препарата.

**Ключевые слова:** биоактивная конформация, лиганд, систематический поиск, молекулярная динамика, торсионное движение, дистанционная геометрия, стереоизомер, дескриптор.

The article provides a comparative description of the programs that are used to create ligand conformations. Based on the analyzed data, the online program FROG2 was selected because of the following characteristics: accessibility, ability to generate conformations for cyclic molecules, stereoisomers, interface convenience, high rate of conformation generation. It allows to get a sufficient number of conformations for ligands, in the definite file formats, which could be input files for the other programs and that allows to calculate descriptors of the chemical structures of conformations. The obtained descriptors are necessary for further evaluation of the biological activity of the compound – a potential active substance in the development of a new medicine.

**Keywords:** bioactive conformation, ligand, systematic search, molecular dynamics, torsion motion, distance geometry, stereoisomer, descriptor.

**Введение.** Как правило, биологическая активность фармацевтической субстанции определяется одной «биоактивной» конформацией её молекул, которую необходимо обнаружить среди всего множества конформаций с минимальными показателями энергии. Для компьютерного дизайна лекарственных препаратов (драг-дизайна) поиск молекул химических соединений с такой конформацией является важным и необходимым, т. к. сведения об активной конформации вещества позволяют конструировать наиболее эффективные биологически активные лиганды для заданной мишени. Принято считать, что биоактивная конформация не всегда имеет минимальную энергию из всех образуемых конформаций, однако она не может превышать по энергии глобальный минимум [1], [2]. Следовательно, определение низкоэнергетических конформаций играет важную роль для понимания и установления корреляции между структурой молекулы и её биологической активностью.

Экспериментальные методы установления структуры химических молекул, например, такие как ЯМР, способны предоставить информацию всего лишь о некоторых конформациях, наиболее часто встречающихся в растворе. Наиболее развернутая картина как и по количеству конформаций, так и об особенностях конформационного пространства может быть получена только с использованием теоретических методов, основы которых нашли свое применение в ряде компьютерных программ. Временные затраты конформационного анализа такими методами зависят от способа расчета энергии полученных конформаций. Энергия может рассчитываться при помощи квантово-механических методов или методов молекулярной механики. Квантово-механические расчеты являются длительными и практически не применимы для больших и подвижных молекулярных систем, поэтому в программах конформационного поиска для оценки показателей энергии чаще всего используются методы молекулярной механики.

Существует много различных методов поиска энергетического минимума молекулярных систем. Все используемые подходы можно систематизировать в виде следующих основных групп:

- 1) систематический поиск – полный анализ конформационного пространства путем систематической проверки всех вариантов конформаций;
- 2) случайный поиск – значения торсионных углов молекулярной модели задаются случайным образом;
- 3) моделирование молекулярной динамики – компьютерное моделирование реальных движений в молекулярной системе во времени;
- 4) эволюционный поиск (генетический алгоритм и стратегия эволюции) – торсионные углы в молекуле изменяются по определенному алгоритму, основанному на генетических принципах.

Систематический поиск дает наилучшие результаты в конформационном анализе. С помощью систематического поиска теоретически можно получить полную картину потенциальной поверхности путем перебора всех возможных величин торсионных углов (вращение различных фрагментов молекулы вокруг всех разрешенных для вращения связей).

Совершенно на другом подходе к исследованию конформационного пространства основан метод Монте-Карло, который встречается под названием «метод случайного поиска». Данный метод имеет стохастическую природу, и при его использовании рассматриваемая конформация изменяется случайным образом, превращаясь из одной в другую. Такой поиск начинается с определения и дальнейшего изменения оптимизированной структуры. Каждый шаг интеграционной процедуры заключается в присваивании молекуле случайным образом новых значений торсионных углов либо декартовых координат, после этого минимизируется энергия каждой из полученных конформаций, и алгоритм повторяется снова. Энергетически минимизированная конформация сравнивается с предыдущей и отображается на выходе лишь в том случае, если она уникальна и величина полной энергии молекулы ниже, чем была на предыдущем шаге. Если значение полной энергии молекулы выше, то отбор получившейся конформации происходит с учетом критерия Метрополиса. Наиболее существенное преимущество метода случайного поиска – возможность изучать молекулы любого размера. Данный метод широко и успешно применим для исследования циклических молекул, поскольку системы, содержащие циклы, систематический поиск анализирует с трудом [1], [2].

На сегодняшнее время широкое применение нашел метод молекулярной динамики на основе принципов молекулярной механики. Цель молекулярной динамики заключается в воспроизведении движения молекулы во времени. Подход молекулярной динамики включает ряд особенностей, например то, что такой метод позволяет преодолевать энергетические барьеры, разделяющие различные конформации в отличие от консервативных процедур оптимизации геометрии. Для увеличения количества конформаций, выбранных в результате моделирования молекулярной динамики, используется повышение температуры моделирования. При такой высокой температуре возрастают шансы получить конформационное разнообразие, т. к. молекулы становятся способными преодолевать даже самые высокие энергетические барьеры, разделяющие некоторые конформации. Основная задача высокотемпературного моделирования – обеспечить молекулу достаточным количеством кинетической энергии для преодоления барьеров между различными конформациями.

Эволюционный поиск основан на применении генетических принципов при расчете изменений торсионных углов в конформационном анализе молекулярной системы. Термины и понятия, используемые в этом методе, соответствуют оригинальным терминам из генетики: индивидуумы, мутации, кроссинговер, воспроизводство. Конформационное пространство исследуется с помощью повторяющихся шагов кроссинговера и мутаций. При этом могут быть найдены не все минимумы, однако, как правило, быстро находится если не глобальный минимум, то большинство локальных минимумов с очень низкой величиной энергии.

Каждый из вышеописанных методов имеет свои преимущества и недостатки. Например, систематический поиск подвержен «комбинаторному взрыву» количества конформаций и не применим к исследованию молекул, содержащих циклы. При использовании методов случайного поиска требуется достаточно времени для подтверждения полноты покрытия конформационного пространства. Эволюционный поиск, основанный на применении генетического ал-

горитма, способен быстро сходиться в близкой окрестности минимума, однако может ошибочно сойтись в локальном минимуме. Для того, чтобы проверить соответствие и правильность методов конформационного поиска используются способы сравнения со стандартными тестовыми наборами. Сравнение осуществляется по числу конформаций, их энергии. Одним из наиболее применяемых способов является сравнение полученных программой конформеров с экспериментально определенными структурами из баз кристаллографических данных [1], [2].

В основной части статьи будут проанализированы программы, которые используются для генерации конформаций в хемоинформатике и биоинформатике, а именно: OMEGA, FROG2, VegaZZ.

### **Основная часть.**

#### **1. Анализ программы генерации конформаций OMEGA.**

Коммерчески доступная программа OMEGA разработана компанией OpenEye Scientific Software Inc. Данная программа представляет собой инструмент для создания трехмерных конформаций молекул. OMEGA предлагает два различных способа генерации конформаций, в основе которых лежит моделирование молекулярной динамики: метод торсионного движения и метод дистанционной геометрии.

Метод торсионного движения предназначен для генерации конформаций молекул, которые не содержат в своей структуре циклов, в то время как метод дистанционной геометрии предназначен для исследования любых молекул, но был разработан конкретно для тех, которые содержат в своей структуре макроциклы.

Проанализировано научное исследование, в ходе которого сравнивали способность разных компьютерных программ для генерации конформаций получать биоактивные конформации (изучалась программа OMEGA 2.0). В проведенном исследовании рассматривали библиотеку из 778 структур молекул лекарственных веществ и фармакологически активных соединений из базы данных Protein Data Bank. Были применены два подхода к генерации конформаций: высокопроизводительный (для работы с большими базами данных) и высококачественный (для изучения центров связывания и повышения качества описания ведущей структуры). Результаты исследования выявили, что качество конформационных моделей определяется выбором оптимального компромисса между плотностью покрытия конформационного пространства и сложностью вычислительного процесса используемого алгоритма. При увеличении размера и подвижности исследуемой молекулы всегда необходимо увеличивать размер конформационного ансамбля для сопоставимого качества представления молекулы. В результате исследования программа OMEGA показала себя с лучшей стороны как генератор высококачественных моделей [1].

Также имеются данные исследований о способности OMEGA в полной мере просматривать конформационное пространство. Для программы была проведена валидация процесса генерации конформаций различных биологически активных соединений. В исследованиях было доказано, что программа успешно генерирует 3D структуры: наблюдали неплохое соответствие между сгенерированными конформациями и кристаллографическими данными, представленными в Кембриджской структурной базе данных (CSD) [3]. Кроме того, показано, что программа OMEGA обеспечивает оптимальный баланс между скоростью и производительностью при анализе большого количества данных [4].

Также было обнаружено, что это самый быстрый из коммерчески доступных генераторов конформаций [5]. Данные о стоимости программы расположены по электронному адресу <https://www.eyesopen.com/omega>. В свободном доступе на английском языке имеется руководство по использованию программы.

#### **2. Анализ программы генерации конформаций FROG2.**

Свободно-доступная онлайн программа FROG2 (от англ. FRee Online drug 3D conformation Generator) предназначена для 3D-генерации конформаций. Интерфейс программы представлен на рисунке 1. Данная программа позволяет генерировать конформационные ансамбли для малых молекул. В качестве входных данных можно задавать 1D или 2D структуру. Программа осуществляет генерацию конформаций, используя двухстадийный подход Монте-Карло в пространстве двугранных углов. После загрузки в программу нужных соединений (лигандов) в заданном формате (SMILES – 1D система молекулярного ввода строки; sdf или mol2 –

2D файлы ввода структурных данных) осуществляется генерация конформаций в 3D формат в форматах mol2, sdf и pdb. Frog2 способна обнаруживать стереоизомеры и в дальнейшем делать выборки конформаций для различных стереоизомеров. Программа также включает в себя новые средства минимизации потенциальной энергии для сгенерированных конформаций и способы генерации конформаций для молекул, содержащих циклы. Оптимизированный генератор конформационных ансамблей FROG2 (в отличие от предыдущей версии программы FROG1) позволяет использовать меньше времени для вычислений. Высокая скорость получения генерируемых конформационных ансамблей делает возможным исследование больших количеств анализируемых соединений с помощью FROG2 [6]. Онлайн-ресурс программы и сопроводительная обучающая документация находятся в свободном доступе по электронному адресу <http://bioserv.rpbs.univ-paris-diderot.fr/Frog2>.

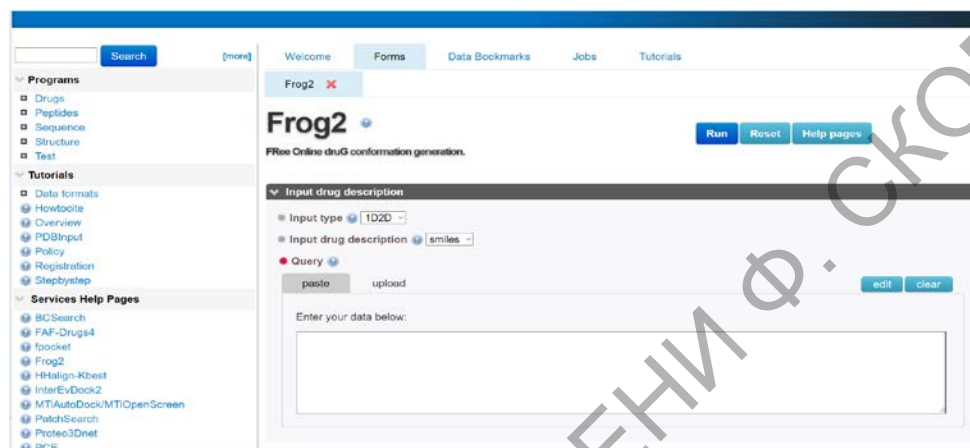


Рисунок 1 – Онлайн-программа Frog2 (окно для ввода данных)

Для генерации конформаций интересующей молекулы (лиганда) необходимо загрузить в программу файл, в котором представлено описание этой молекулы (лиганда) в определенном, предусмотренном программой, формате. После загрузки файла следует выбрать параметры, согласно которым будет выдан полученный результат. Параметры задаются в интерфейсе программы, представленном на рисунке 2.

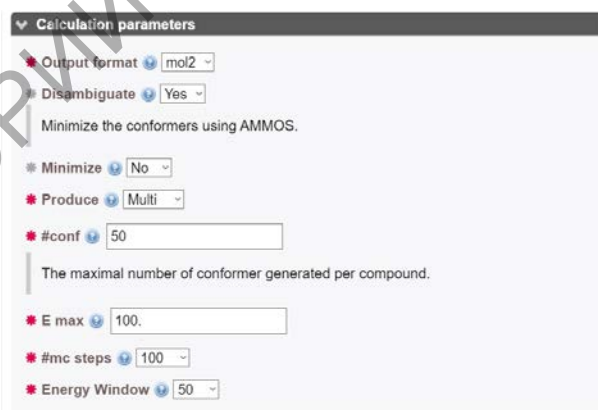


Рисунок 2 – Параметры для получаемых данных

В таблице параметров указывается формат данных для отображения искоемых конформации (Output format), количество шагов интеграционной процедуры метода Монте-Карло (#mc steps), величина «энергетического окна» (Energy window), максимальное количество конформаций для генерации (#conf), а также выбрать возможность минимизировать энергию конформаций (Minimize) и возможность получения мультиконформеров (Produce).

После загрузки исходных данных и задания параметров поиска конформаций, запускается генерация конформеров. Результат работы программы представлен на рисунке 3.

https://mobyli.rpbs.univ-paris-diderot.fr/data/jobs/Frog2/15091503460884

get help back to form remove job download

results

Results

Unambiguous smiles (Drug\_3DStructure)

Frog.smiles (smiles)

```
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
```

mol1\_0  
mol1\_1  
mol1\_2  
mol1\_3

full screen bookmark as Frog.smiles or further analysis

PDB file (Drug\_3DStructure)

Frog.pdb (PDB)

MODEL	1
CDPNO	mol1_1_1
AUTHOR	GENERATED BY OPEN BABEL 2.3.0
ATOM	1 C1 <O> A 1 -0.785 -5.779 0.369 1.00 0.00 C
ATOM	2 S1 <O> A 1 0.825 -6.248 0.982 1.00 0.00 S
ATOM	3 C2 <O> A 1 1.274 -7.726 0.103 1.00 0.00 C
ATOM	4 C3 <O> A 1 1.429 -7.547 -1.423 1.00 0.00 C
ATOM	5 C4 <O> A 1 0.674 -6.348 -2.082 1.00 0.00 C
ATOM	6 S2 <O> A 1 -0.975 -6.242 -1.348 1.00 0.00 S

full screen NGL PV bookmark as Frog.pdb or further analysis

Logs (Text)

Frog.log

```
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
[C]1[S]1([C@]([C@]([C]S1)[C])C)C2=C([C]([C]([C]C2)C)3[C]([C]([C]S[C]3)
72 conformation(s) calculated. For some isomers 0 have bad energies.
```

full screen bookmark as Frog.log or further analysis

Рисунок 3 – Вариант представления выводных данных

Программа представила конформации молекулы в запрашиваемом формате. Указано количество сгенерированных конформаций, а также количество изомеров среди них. Согласно литературным данным, программа Frog2 является самой быстрой по скорости генерации из предложенных в свободном бесплатном доступе программ [7].

Баланс между скоростью генерации конформеров и конформационным разнообразием достигается в зависимости от целей проводимого исследования и количества загружаемых в программу соединений. Отключение второго этапа метода Монте-Карло, которое можно выбрать в настройках на сервере, приведет к более быстрой генерации, но количество полученных конформаций будет меньше. По средним расчетам Frog2 находит биоактивную конформацию в пределах RMSD (среднеквадратичное отклонение)  $0,73 \pm 0,42 \text{ \AA}$ . Примеры, демонстрирующие конформационные ансамбли, полученные Frog2 и коммерчески доступной программой Omega, для двух молекул набора данных Astex, можно увидеть на рисунке 4. На данном рисунке слева представлены конформации, сгенерированные Frog2, справа – программой Omega. В части а) рисунка отображены конформации для лиганда с идентификатором PDB (Protein Data Bank) 1jla, в части b) рисунка показаны конформации для лиганда с идентификатором PDB 1meh. Также в каждой из частей рисунка вверху отображены структурные формулы лигандов, для которых проводился конформационный поиск. Полученные с использованием Frog2 и Omega результаты демонстрируют небольшое превосходство Omega (которая является коммерческой программой) со значением показателя среднеквадратичного отклонения равным  $0,69 \pm 0,37 \text{ \AA}$ . Согласно вычисленным значениям RMSD и визуальному анализу, можно сделать вывод, что и Frog2, и Omega довольно хорошо исследуют конформационное пространство и способны генерировать конформации, близкие к биоактивной. Кроме того, у пользователя есть возможность минимизировать энергию генерируемых конформеров с помощью Frog2 или минимизировать собственную библиотеку соединений в 3D, задавая параметры выходных данных. Также в литературных источниках отмечено, что точность генерации конформаций в программе Frog2 довольно высока. В результате сравнения сгенерированных конформаций (было использовано 10000 случайно выбранных соединений) с данными, хранящимися в кристаллографической базе данных химических соединений ChemBridge diverset (<http://www.chembridge.com>) Frog2 смогла сгенерировать конформеры для более чем 98 % соединений, что говорит об адекватности и надежности конформационного поиска с использованием данной программы. Из минусов Frog2 можно отметить, что иногда возникают ошибки при генерации стереоизомеров для циклических соединений и появляются неточности относительно создания конформаций для определенных химических групп [6], [7].

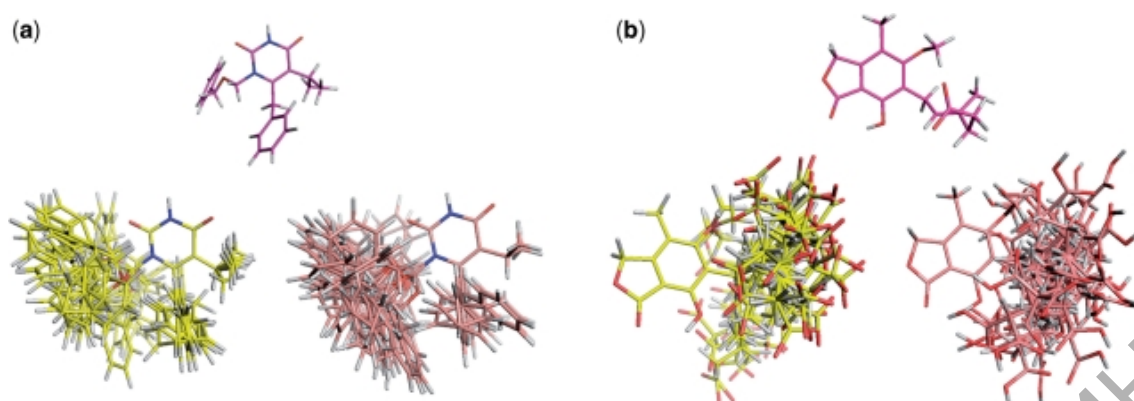


Рисунок 4 – Конформационные ансамбли для двух малых молекул базы данных Astex, сгенерированные Frog2 и Omega

### 3. Анализ программы генерации конформаций VegaZZ.

Программное обеспечение VegaZZ для Windows имеется в свободном доступе и загружается на веб-сайте <https://www.ddl.unimi.it/>. Пользователь может проводить все имеющиеся операции по молекулярному моделированию на персональном компьютере. Интерфейс данной программы представлен на рисунке 5. Двухмерный молекулярный редактор Ketcher позволяет рисовать молекулы, а последующая команда «Отправить в VegaZZ» предоставляет полученную структуру трехмерной модели.

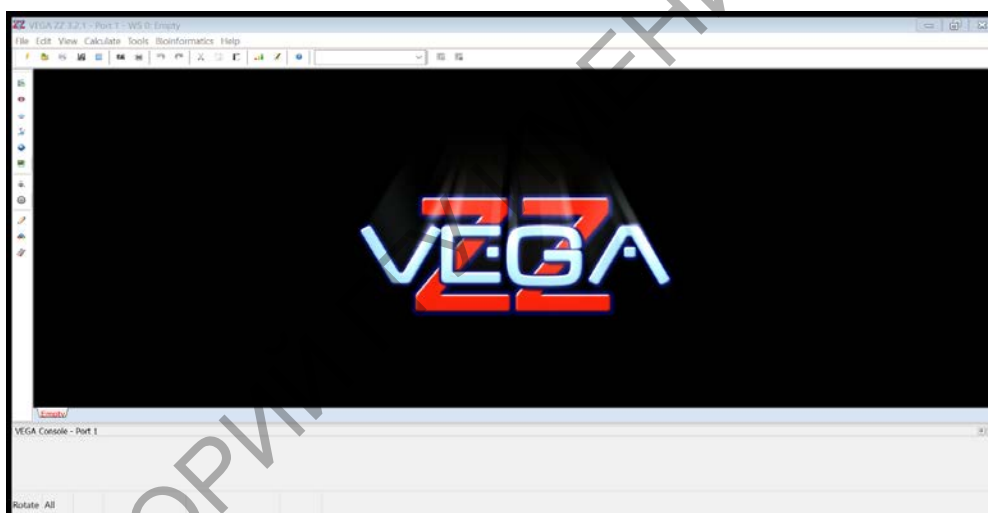


Рисунок 5 – Интерфейс программы VegaZZ

Конформационный анализ выполняется с помощью встроенной программы АММР с силовым полем SP4. Этот метод определен как систематический, с возможностью минимизировать энергии всех конформаций, остальные параметры могут устанавливаться по умолчанию либо выбираться пользователем. Перед запуском расчета необходимо задать количество шагов изменения угла и сохранить выходные файлы (траектория, выход и энергия). Чтобы проанализировать результаты, необходимо нажать «Расчитать, анализ», затем открыть файл в формате .dcd и создать график, выбирая опцию «График энергии». Наиболее стабильную конформацию можно получить при указании на пользовательском интерфейсе диаграммы с наименьшей энергией. В программе для просмотра будут доступны и другие полученные для данной молекулы конформации. Для каждой конформации будет представлена энергия на графике. Пример такого графика изображен на рисунке 6.

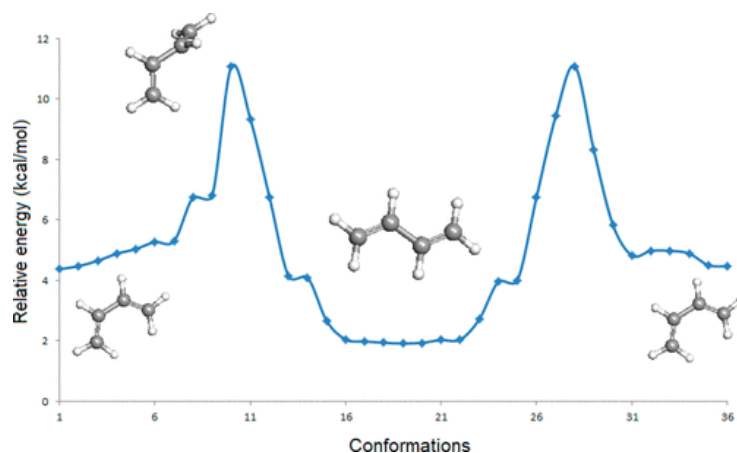


Рисунок 6 – Конформации, сгенерированные в программе и соответствующая им потенциальная энергия

Однако нет данных о том, что программа VEGA ZZ учитывает количество стереоизомеров при генерации конформаций, также в литературных источниках нет подтверждающей информации о том, насколько точно и правильно сгенерированы конформации. Программа показала хорошую скорость генерации конформаций для простых химических молекул (в литературном источнике была рассмотрена генерация конформаций для *n*-бутана, циклического ацетала, аминокислот и многоатомных спиртов). Следует отметить, что программа способна генерировать конформации для молекул, содержащих циклы, молекул, содержащих внутримолекулярную водородную связь и молекул, для которых характерен мезомерный эффект. Также VEGA ZZ позволяет рассчитывать значения заданных торсионных углов для каждой молекулы. В целом программа предназначена для генерации конформеров низкомолекулярных соединений, однако нет информации о том, насколько хорошо она будет генерировать биоактивные конформации для лигандов – потенциальных лекарственных препаратов [8].

**Заключение.** После анализа информации о программах для использования в качестве наиболее оптимальной можно рассматривать программу для генерации конформаций FROG2 ввиду ее следующих характеристик: доступности, способности генерировать конформации для циклических молекул, стереоизомеров, удобства интерфейса; высокой скорости генерации конформаций. Также, исходя из данных литературных источников, программа FROG2 продемонстрировала приемлемые и сопоставимые характеристики (RMSD) при её сравнении с коммерчески доступной программой OMEGA.

## Литература

1. Молекулярное моделирование : теория и практика / Х.-Д. Хельтзе [и др.]. – М. : БИНОМ. Лаборатория знаний, 2010. – 318 с.
2. Иванов, А. С. Основные принципы молекулярного конформационного анализа для медико-биологов / А. С. Иванов // Биомедицинская химия. – 2007. – Т. 53, Вып. 6. – С. 713–728.
3. Hawkins, P. C. D. Conformer Generation with OMEGA : Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database / P. C. D. Hawkins, A. G. Skillman, G. L. Warren [et al.] // Chem. Inf. Model. – 2010. – Vol. 50. – P. 572–584.
4. Perola, E. Conformational Analysis of Drug-like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding / E. Perola, P. S. Charifson // Med. Chem. – 2004. – Vol. 47. – P. 2499–2510.
5. Friedrich, N.-O. de Bruyn Kops. Benchmarking Conformer Ensemble Generators / N.-O. de Bruyn Kops Friedrich, F. Fachsenberg, K. Sommer, M. Rarey, J. Kirchmair // Chem. Inf. Model. – 2017. – Vol. 5. – P. 2719–2728.
6. Miteva, M. Frog2 : Efficient 3D conformation ensemble generator for small compounds / M. Miteva, F. Guyon, P. Tuffery // Nucleic Acids Research. – 2010. – Vol. 38. – P. W622–W627.
7. Ebejer, J.-P. Freely Available Conformer Generation Methods : How Good Are They? / J.-P. Ebejer, G. M. Morris, Ch. M. Deane // Chem. Inf. Model. – 2012. – Vol. 52. – P. 1146–1158.
8. Pedretti, A. VEGA : a versatile program to convert, handle and visualize molecular structure on Windows-based PCs / A. Pedretti, L. Villa, G. Vistoli // Journal of Molecular Graphics and Modelling. – 2002. – Vol. 21. – P. 47–49.