

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

СОГЛАСОВАНО
Генеральный директор
ЗАО «АйТи»


Бакиев О.Р.
“22” сентября 2011 г.

УТВЕРЖДАЮ
Ректор НИУ ИТМО


Васильев В.Н.
“22” сентября 2011 г.

МНОГОПРОФИЛЬНАЯ ИНСТРУМЕНТАЛЬНО-
ТЕХНОЛОГИЧЕСКАЯ ПЛАТФОРМА СОЗДАНИЯ
И УПРАВЛЕНИЯ РАСПРЕДЕЛЕННОЙ СРЕДОЙ
ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ CLAVIRE

ПРИКЛАДНОЙ СЕРВИС ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СКРИНИНГА И
МОДЕЛИРОВАНИЯ ФАРМАКОЛОГИЧЕСКОЙ АКТИВНОСТИ

ОПИСАНИЕ ПРОГРАММЫ

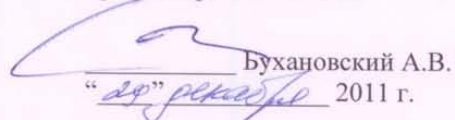
ЛИСТ УТВЕРЖДЕНИЯ

RU.СНАБ.80066-06 13 52-ЛУ

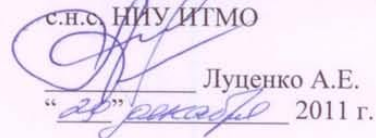
Ине.№ подл.	Подп. и дата
Взам.инв.№	Подп. и дата
Ине.№ дубл.	Подп. и дата
Подп. и дата	Подп. и дата

Представители
Организации-разработчика

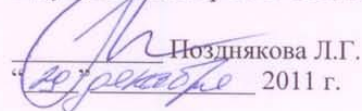
Руководитель разработки,
профессор НИУ ИТМО


Бухановский А.В.
“22” сентября 2011 г.

Ответственный исполнитель,
с.н.с. НИУ ИТМО


Луценко А.Е.
“22” сентября 2011 г.

Нормоконтролер
ведущий инженер НИУ ИТМО


Позднякова Л.Г.
“22” сентября 2011 г.

2011

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**УТВЕРЖДЕН
RU.СНАБ.80066-06 13 52-ЛУ**

**МНОГОПРОФИЛЬНАЯ ИНСТРУМЕНТАЛЬНО-
ТЕХНОЛОГИЧЕСКАЯ ПЛАТФОРМА СОЗДАНИЯ
И УПРАВЛЕНИЯ РАСПРЕДЕЛЕННОЙ СРЕДОЙ
ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ CLAVIRE**

**ПРИКЛАДНОЙ СЕРВИС ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СКРИНИНГА И
МОДЕЛИРОВАНИЯ ФАРМАКОЛОГИЧЕСКОЙ АКТИВНОСТИ**

ОПИСАНИЕ ПРОГРАММЫ

RU.СНАБ.80066-06 13 52

ЛИСТОВ 27

Ине.№ подл.	
Подп. и дата	
Взам. ине.№	
Ине.№ дубл.	
Подп. и дата	

АННОТАЦИЯ

Документ содержит описание прикладного сервиса высокопроизводительного скрининга и моделирования фармакологической активности. Прикладной сервис реализует композитное приложение для многоцелевой инструментально-технологической платформы (МИТП) CLAVIRE, которое позволяет, основываясь на заданной структуре рецептора, построить модель взаимодействия с ним каждого из кандидатов, отобрать наилучших и оценить энергию макромолекулярной системы высокоточными методами. Прикладной сервис разработан в ходе выполнения проекта «Создание распределенной вычислительной среды на базе облачной архитектуры для построения и эксплуатации высокопроизводительных композитных приложений» (Договор № 21057 от 15 июля 2010 г., шифр 2010-218-01-209) в рамках реализации постановления Правительства РФ № 218 «О мерах государственной поддержки развития кооперации российских высших учебных заведений и организаций, реализующих комплексные проекты по созданию высокотехнологичного производства».

СОДЕРЖАНИЕ

1.	ОБЩИЕ СВЕДЕНИЯ	4
2.	ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ	5
2.1.	Область применения	5
2.2.	Функциональное назначение.....	5
2.3.	Ограничения на применение	5
3.	ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ	6
3.1.	Структура композитного приложения	6
3.2.	Характеристики модулей композитного приложения	9
3.3.	Описание модулей композитного приложения в МИТП	15
4.	ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА	22
5.	ВЫЗОВ И ЗАГРУЗКА.....	22
6.	ВХОДНЫЕ ДАННЫЕ	23
7.	ВЫХОДНЫЕ ДАННЫЕ	23
	ПЕРЕЧЕНЬ СОКРАЩЕНИЙ	25
	ПЕРЕЧЕНЬ ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	26

1. ОБЩИЕ СВЕДЕНИЯ

Прикладной сервис (ПС) высокопроизводительного скрининга и моделирования фармакологической активности RU.СНАБ.80066-06 01 52 реализует композитное приложение, которое позволяет, основываясь на заданной структуре рецептора, построить модель взаимодействия с ним каждого из кандидатов, отобрать наилучших и оценить энергию макромолекулярной системы высокоточными методами. ПС разработан в ходе выполнения проекта «Создание распределенной вычислительной среды на базе облачной архитектуры для построения и эксплуатации высокопроизводительных композитных приложений» (Договор № 21057 от 15 июля 2010 г., шифр 2010-218-01-209) в рамках реализации постановления Правительства РФ № 218 «О мерах государственной поддержки развития кооперации российских высших учебных заведений и организаций, реализующих комплексные проекты по созданию высокотехнологичного производства».

ПС функционирует в рамках распределенной среды облачных вычислений под управлением многофункциональной инструментально-технологической платформы (МИТП) CLAVIRE RU.СНАБ.80066-06. Он разработан на предметно-ориентированном языке EasyFlow описания композитных приложений.

ПС использует следующие пакеты прикладных программ, доступные в распределенной среде под управлением МИТП (см. также раздел 3.2):

Autodock Vina - прикладной пакет для выполнение задачи докинга;

NWChem - прикладной пакет для выполнение расчета в рамках подхода QM/MM;

MGL Tools - прикладной пакет для выполнение операции подготовки структуры рецептора;

Q region selector - прикладной пакет для вычисления границы области моделирования квантово-механическими методами;

Sору pack для - вспомогательный скрипт передачи файлов содержащих структуру лигандов из базы данных;

Перечисленные пакеты описываются на языке EasyFlow и регистрируются в базе пакетов МИТП.

2. ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ

2.1. Область применения

Прикладной сервис высокопроизводительного скрининга осуществляет автоматическое выявление и отбор соединений с заданной фармакологической активностью с последующим моделированием взаимодействия с мишенью высокоточными методами. Высокая производительность средств докинга позволяет быстро и эффективно отсортировать лучших кандидатов по аффинитету. Применение композиции молекулярной динамики с высокоточными методами первопринципных расчетов позволяет добиваться высокой точности в моделировании встраивания отобранных кандидатов. Интеграция под управлением платформы модулей настроенных под данное композитное приложение, а также наличие собственных модулей, разработанных для данного композитного приложения, дает в руки пользователя высокоэффективный инструмент решения различных биохимических и фармакологических задач, абстрагируя его при этом не только от работы отдельно взятых модулей, но и от решаемых ими наукоемких задач.

2.2. Функциональное назначение

Данный ПС является мощным средством решения следующих задач:

- Обоснование активности существующих препаратов на молекулярном уровне с высокой точностью;
- Проектирование новых фармакологических препаратов высокоточными методами;
- Выявление заданной новой фармакологической активности у существующих лекарственных средств высокоточными методами.

2.3. Ограничения на применение

Как правило стандартные современные средства докинга плохо справляются с задачей в тех случаях, когда в сайте белка наличествуют атомы металлов. Поскольку в данном сервисе на конечном этапе приложения выполняется квантово-химический расчет первопринципными методами, который с легкостью справляется с моделированием металлоорганических структур, данный сервис может быть рассмотрен, как имеющий более широкое применение в рамках докинга белковых систем.

3. ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ

3.1. Структура композитного приложения

На первом этапе интересующая пользователя структура белка-мишени предоставляется им в систему и подготавливается для докинга при помощи пакета MGL Tools. После этого запускается процесс скрининга лигандов, при котором происходит докинг каждого лиганда из базы данных в указанный пользователем сайт белка. Лиганды передаются из базы данных для процесса скрининга посредством скриптового приложения Cору pack. При этом осуществляется двухъярусная параллельность: во-первых, докинг набора лигандов выполняется одновременно, во-вторых, расчет каждого соединения в отдельности осуществляется в несколько параллельных потоков. Моделирование осуществляется программой Autodock Vina. Для докинга каждого отдельно взятого лиганда программа автоматически определяет количество параллельных потоков на ресурсе, где осуществляется расчет. Для каждого соединения определяется несколько вариантов геометрии его встраивания в мишень и соответствующий аффинитет. Далее отбираются лучшие кандидаты и передаются на второй этап.

Второй этап начинается с работы препроцессора Q region selector, который был написан специально для данного композитного приложения. Для каждого лиганда с высоким аффинитетом препроцессор определяет окружающие его атомы белка и содержащие их аминокислотные остатки. Данные остатки вместе с лигандом формируют область для высокоточного расчета. Остальная часть белковой глобулы будет моделироваться средствами молекулярной динамики. Далее запускается основной расчет, который решает задачу оптимизации геометрии для всей системы в целом. При этом система моделируется гибридным методом QM/MM. В основе подхода QM/MM лежит разделение системы на две части: классическую и квантовую. При этом классическая часть описывается методами молекулярной механики, а квантовая – с применением того или иного квантово-химического подхода. Влияние классической подсистемы на квантовую рассматривается как внешнее электростатическое поле. В данной реализации решение этой задачи осуществляется в два этапа. Первый этап – подготовительный. В ходе его осуществляется: (1) задание величин и координат точечных зарядов, моделирующих влияние классической части системы на квантовую и (2) задание корректных граничных условий для квантовой области. Определение величин и координат точечных зарядов определяются типом и координатами атомов, входящих в классическую область в рамках той модели потенциального поля, которая используется

для описания классической подсистемы. Проблема задания граничных условий возникает в тех случаях, когда граница между квантовой и классической областями приходится на валентную связь между атомами. В данной реализации задание корректных граничных условий осуществляется путем дополнения свободной валентности атомами водорода.

После выполнения подготовительного этапа осуществляется оптимизация геометрии системы в рамках QM/MM подхода. Процесс осуществляется итерационно. На каждой итерации производятся: (1) несколько шагов оптимизации геометрии квантовой подсистемы и (2) несколько шагов оптимизации геометрии классической подсистемы. Критерием завершения процесса является малость элементов вектора градиента для классической и квантовой подсистем. При оптимизации геометрии квантовой подсистемы в качестве квантово-химического метода выбирается метод DFT B3LYP в базисе умеренного размера (6-31G). Решается обычная квантово-химическая задача, но с учетом внешнего электростатического поля, создаваемого точечными зарядами, моделирующими влияние классической подсистемы. Положение этих точечных зарядов корректируется на каждой итерации после выполнения шагов оптимизации классической подсистемы.

Результатом работы композитного приложения является структура смоделированного комплекса макромолекулярного белка-мишени и выбранного лиганда.

Для реализации композитного приложения в соответствии со структурой на рис. 3.1, используются следующие прикладные пакеты:

MGL Tools - прикладной пакет для выполнения операции подготовки структуры рецептора;

Autodock Vina - прикладной пакет для выполнения задачи докинга;

Q region selector - прикладной пакет для вычисления границы области моделирования квантово-механическими методами;

NWChem - прикладной пакет для выполнения расчета в рамках подхода QM/MM;

В листинге 3.1 приведен скрипт композитного приложения на языке EasyFlow, а на рис. 3.1 - результат его интерпретации в МИТП, который создает очередь из пяти последовательных процессов запуска пакетов в виде AWF.

Листинг 3.1 Описание композитного приложения на языке EasyFlow

```
require receptor_pdb;  
  
step ligand_produce runs copy_pack  
();  
  
step convert runs MGL_Tools  
(  
    receptor_pdb = sweep ligand_produce.Result.ligands
```



```

);

step Docking runs vina
(
  receptor = sweep convert.Result.outs["receptor.pdbqt"],
  ligand = ligand_pdbqt,
  search_space_center_x = 15,
  search_space_center_y = 8.5,
  search_space_center_z = 2,
  search_space_size_x = 6.0,
  search_space_size_y = 6.0,
  search_space_size_z = 4
);

step NWChem_prepare runs q_region_selector
(
  receptor = receptor_pdb,
  ligand_confs = Docking.Result.outs["ligand_out.pdbqt"]
);

step QMMM runs NWChem
(
  config_input      = sweep NWChem_prepare.Result.models,
  coordinates_input = sweep NWChem_prepare.Result.configs,
);

```

The screenshot displays a workflow editor window. On the left, a script is written in a code editor. On the right, a flowchart visualizes the execution sequence of the script's steps. The script includes steps for ligand production, conversion, docking, preparation of a region selector, and finally QMMM simulation. The flowchart shows a vertical sequence of boxes connected by dashed arrows, representing the data flow between these steps.

```

1 require receptor_pdb;
2
3 step ligand_produce runs copy_pack
4 ();
5
6 step convert runs MGL_Tools
7 (
8   receptor_pdb = sweep ligand_produce.Result.ligands
9 );
10
11 step Docking runs vina
12 (
13   receptor = sweep convert.Result.outs["receptor.pdbqt"],
14   ligand = ligand_pdbqt,
15   search_space_center_x = 15,
16   search_space_center_y = 8.5,
17   search_space_center_z = 2,
18   search_space_size_x = 6.0,
19   search_space_size_y = 6.0,
20   search_space_size_z = 4
21 );
22
23 step NWChem_prepare runs q_region_selector
24 (
25   receptor = receptor_pdb,
26   ligand_confs = Docking.Result.outs["ligand_out.pdbqt"]
27 );
28
29 step QMMM runs NWChem
30 (
31   config_input      = sweep NWChem_prepare.Result.models,
32   coordinates_input = sweep NWChem_prepare.Result.configs,
33 );

```

The flowchart consists of the following steps in a vertical sequence:

- ligand_produce** (copy_pack)
- convert** (MGL_Tools)
- Docking** (vina)
- NWChem_prepare** (q_region_selector)
- QMMM** (NWChem)

Рис. 3.1 - Подготовка композитного приложения к запуску на выполнение в МИТП

3.2. Характеристики модулей композитного приложения

В ходе процесса исполнения композитного приложения осуществляется интеграция вычислительных ресурсов под управлением МИТП-Ц для исполнения пакетов MGL Tools, Autodock Vina, Q region selector и NWChem. Композитное приложение описывает поток, заданий на языке EasyFlow, интерпретируемый в МИТП-Ц CLAVIRE.

3.2.1 Пакет MGL Tools

3.2.1.1 Общие сведения о пакете MGL Tools

Прикладной пакет MGL Tools является программным комплексом сторонней разработки, предназначенным для работы с молекулярными структурами, их обработки конвертирования и визуализации [1]. Программный комплекс MGL Tools разработан на языке Python. Комплекс MGL Tools функционирует на ЭВМ под управлением Unix-подобных и Windows ОС.

3.2.1.2 Используемые технические средства

Требования к аппаратной совместимости пакета MGL Tools: персональный компьютер или узел кластера на базе процессора не менее Pentium IV, свободное дисковое пространство – не менее 10 ГБ, объем оперативной памяти – не менее 256 МБ.

Требования к информационной и программной совместимости: программа предназначена для функционирования на персональном компьютере с операционными системами Unix типа или серии Windows. На компьютере должно быть установлено приложение python 2.5.

3.2.1.3 Вызов и загрузка

Для вызова пакета MGL Tools требуется перейти в директорию в которой находится приложение и запустить скрипт. Пример:

```
MGLTools Входной_файл.inp > Выходной_файл.out.
```

В случае удачного запуска и отработки приложения в директории появится отображение выходных данных в запрашиваемом формате. В случае удачного запуска и отработки приложения в директории появится файл: Выходной_файл.out. с выходными данными.

3.2.1.4. Входные и выходные данные

Входным параметром пакета является указанный в командной строке или скрипте запуска входной файл.

Результатом работы пакета являются выходные файлы указанного в скрипте запуска формата. Пример:

```
prepare_ligand4.py -l имя_файла
```

В случае удачного запуска и отработки приложения в директории появится файл: Выходной_файл.pdbqt с данными.

3.2.2 Пакет молекулярного моделирования Autodock Vina

3.2.2.1 Общие сведения о пакете Autodock Vina

Прикладной пакет Autodock Vina является программным комплексом сторонней разработки, предназначенным для выполнения задачи докинга молекулярных структур [2]. Приложение Autodock Vina функционирует на ЭВМ под управлением Unix-подобных Windows и Mac OS.

3.2.2.2 Функциональное назначение пакета

Программный комплекс Autodock Vina предназначен для выполнения задачи докинга молекулярных структур.

3.2.2.3. Используемые технические средства

Требования к аппаратной совместимости пакета Autodock Vina: персональный компьютер или узел кластера на базе процессора не менее Pentium IV, свободное дисковое пространство – не менее 10 ГБ, объем оперативной памяти – не менее 256 МБ.

Требования к информационной и программной совместимости: программа предназначена для функционирования на персональном компьютере с операционными системами Unix типа или серии Windows.

3.2.2.4. Вызов и загрузка

Для вызова пакета Autodock Vina требуется в консольном приложении перейти в директорию, в которой находится приложение, и запустить исполняемый файл.

В случае удачного запуска и отработки приложения в директории появится отображение выходных данных в запрашиваемом формате. Пример:

```
vina.exe config --config конфигурационный_файл.txt --log log.txt входной_файл.inp >
выходной_файл.out.
```

В случае удачного запуска и отработки приложения в директории появится файл Выходной_файл_out.pdbqt с выходными данными и файл log.txt.

3.2.2.5 Входные данные

Входными параметрами пакета являются:

Файл, содержащий структуру рецептора, в формате pdbqt, файл содержащий структуру лиганда, в формате pdbqt и конфигурационный файл, содержащий названия файлов, содержащих структуру рецептора и лиганда соответственно (табл. 3.1 и 3.2).

Таблица 3.1

Входные параметры для Autodock Vina. Перечень входных файлов

Имя параметра	Описание параметра
имя_файла. Pdbqt	pdbqt-файл, содержащий структуру рецептора в формате pdbqt
имя_файла. Pdbqt	pdbqt-файл, содержащий структуру лиганда в формате pdbqt
имя_конфигурационного файла	txt-конфигурационный файл

Таблица 3.2

Входные параметры для Autodock Vina. Перечень входных параметров для конфигурационного файла.

Имя параметра	Описание параметра
Ligand	Имя pdb-файла, содержащего структуру лиганда
receptor	Имя pdb-файла, содержащего структуру рецептора
search_space_center_x	Центр области поиска молекулярного сайта по координате x
search_space_center_y	Центр области поиска молекулярного сайта по координате y
search_space_center_z	Центр области поиска молекулярного сайта по координате z
search_space_size_x	Длина ребра параллелепипеда, соответствующего области поиска по координате x
search_space_size_y	Длина ребра параллелепипеда, соответствующего области поиска по координате y
search_space_size_z	Длина ребра параллелепипеда, соответствующего области поиска по координате z
Receptor	pdb-файл, содержащий структуру рецептора

3.2.2.6. Выходные данные

Результатом работы пакета являются выходной файл log.txt, содержащий информацию о ходе выполнения расчета, и файл, содержащий структуры с наибольшим аффинитетом. В случае удачного запуска и отработки приложения в рабочей директории появится файл имя_файл_out.pdbqt с выходными данными и файл log.txt.

3.2.2 *Пакет молекулярного моделирования Q region selector*

3.2.2.1 Общие сведения о пакете Q region selector

Программное приложение Q region selector является утилитой, специально разработанной для композитного приложения высокопроизводительного скрининга в коллективе НИИ НКТ СПбГУИТМО, предназначенной для обработки результатов докинга, отбора структур с максимальным аффинитетом, преобразования формата pdb в pdbqt, вычисления границы области моделирования квантово-механическими методами. Приложение Q region selector функционирует на ЭВМ под управлением Unix-подобных ОС.

3.2.2.2 Функциональное назначение пакета

Программное приложение Q region selector предназначен для выполнения следующих задач:

- обработки результатов докинга и отбора структур с максимальным аффинитетом;
- преобразования формата pdb в pdbqt;
- вычисления границы области моделирования квантово-механическими методами.

3.2.2.3. Используемые технические средства

Требования к аппаратной совместимости пакета Q_region_selector: персональный компьютер или узел кластера на базе процессора не менее Pentium IV, свободное дисковое пространство – не менее 1 Гб, объем оперативной памяти – не менее 256 Мб.

Требования к информационной и программной совместимости: программа предназначена для функционирования на персональном компьютере с операционными системами Unix-типа.

3.2.2.4. Вызов и загрузка

Для вызова пакета Q region selector требуется в консольном приложении перейти в директорию, в которой находится приложение, и запустить исполняемый файл:

Файл_лиганда.pdbqt файл_рецептора.pdb

В случае удачного запуска и отработки приложения в текущем каталоге появится набор файлов с расширениями файл.nw и файл.pdb.

3.2.2.5. Входные и выходные данные

Входными параметрами пакета являются:

Файлы, содержащие структуру рецептора в формате .pdbqt,

Файл, содержащий структуру лиганда в формате .pdb (табл. 1.2).

Результатом работы пакета являются набор файлов, содержащих координаты атомов молекулярных комплексов с наибольшим аффинитетом и комплект соответствующих входных файлов для программы NWChem с указанием квантовой области.

3.2.3 Прикладной пакет квантовой химии NWChem

3.2.3.1 Общие сведения о пакете NWChem

Прикладной пакет NWChem является программным компонентом, позволяющим производить высокоточные расчеты атомно-молекулярных систем, рассчитывать возбужденные состояния. Пакет содержит методы, базирующиеся на методах Хартри–Фока, функционала плотности, многоконfigurационного самосогласованного поля.

Дистрибутивы пакета существуют как для семейства операционных систем Windows, так и для Linux.

3.2.3.2 Функциональное назначение пакета

Пакет NWChem реализует набор высокоточных методов квантовой химии [3], позволяющих моделировать основное и возбужденные состояния квантовых и молекулярных систем, учитывать их взаимное влияние, оценивать эффекты межэлектронной корреляции, оптимизировать геометрию. Кроме того, пакет реализует метод молекулярной динамики и совмещенный расчет методами молекулярной динамики и квантовой химии.

3.2.3.3 Используемые технические средства

Пакет NWChem может выполняться на программно-аппаратном комплексе, к которому предъявляются следующие требования.

Требования к аппаратной совместимости: персональный компьютер или узел кластера на базе процессора не менее Pentium IV, свободное дисковое пространство – не менее 500 МБ, объем оперативной памяти – не менее 512 МБ.

Требования к информационной и программной совместимости: программа предназначена для функционирования на персональном компьютере в среде ОС Windows, Linux или Mac OS.

3.2.3.4 Вызов и загрузка

Для вызова приложения требуется в консольном приложении перейти в директорию, в которой находится приложение, и выполнить команду:

```
NWChem Входной_файл.nw > Выходной_файл.out
```

В случае удачного запуска и отработки приложения в директории появится файл Выходной_файл.out. с выходными данными.

3.2.3.5 Входные данные

Входным параметром пакета является конфигурационный файл, содержащий параметры задачи и координаты атомов рассчитываемой системы, либо имя файла, содержащего рассчитываемую структуру в формате pdb.

Листинг 3.2.3.5. Пример входного NWChem

```

echo
memory total 800 mb
start asa_am_prep
prepare
  system asa_am_ref
  source asa_am.pdb
  new_top new_seq
  new_rst
  modify segment 4 quantum
  modify atom 2:_OG quantum
  modify atom 2:_HG quantum
  update lists
  ignore
  write asa_am_ref.rst
  write asa_am_ref.pdb # write out PDB file to check structure
end
task prepare

#-----
# this can be replaced by
# task shell "cp asa_ref.rst asa_qmmm.rst"
#-----
prepare
read asa_am_ref.rst
write asa_am_qmmm.rst
end
task prepare
#-----
# noshake solute always have tobe present
# if optimizing classical solvent
#-----
md
  system asa_am_qmmm
  # noshake solute
  # cells 2 2 2
end

basis "ao basis"
  * library 6-31G
end

dft
  print low
  iterations 500
end

#-----
# optimization will be done for three
# regions (one after another)
# esp charges will be used where possible
# to speed up calculations
#-----
qmmm
region qmlink mm_solute
maxiter 5 5
ncycles 200
density espfit
end

```

```
task qmmm dft optimize
```

3.2.3.6 Выходные данные

Результатом работы пакета является текстовый файл, состоящий из блоков, каждый из которых включает группу связанных данных. Данные, получающиеся в результате расчета, можно классифицировать по решаемой задаче. Ниже перечислены основные задачи, для которых выделяются специфические группы выходных параметров:

- расчет основного состояния;
- расчет возбужденного состояния;
- оптимизация геометрии;
- расчет геометрии в седловой точке;
- расчет колебательных частот;
- молекулярная динамика;
- совмещенный расчет методами молекулярной динамики и квантовой химии.

3.3. Описание модулей композитного приложения в МИТП

3.3.1 Платформенное описание прикладного пакета MGL Tools как модуля композитного приложения

Для интеграции пакета MGL Tools в МИТП необходимо описать его взаимодействие с платформой, опираясь на формат представления входных и выходных данных. Платформенный скрипт описания пакета MGL Tools позволяет определить уровень абстракции и интерпретации в работе с входными и выходными параметрами, что обеспечивает гибкость и упрощение процедуры взаимодействия пользователя с платформой на этапе запуска задания. Тем самым необходимость в определении на входе тех или иных параметров по умолчанию устраняется, а сложные составные входные данные декомпозируются на более мелкие составляющие. Это приводит к более эффективному использованию пакета в рамках состава компонентов выполняемого сервиса.

Платформенный скрипт описания пакета MGL Tools определяется в соответствии с листингом 3.3.1.

Листинг 3.3.1. Основная часть параметров описания пакета MGL Tools

```
name "MGL_Tools"  
display_as "MGL_Tools"  
vendor "Iowa State University"
```



```
url "http://onlinelibrary.wiley.com/doi/10.1002/jcc.21334/abstract"  
license "GPLV3"  
description "Very useful and important package"  
  
inputs {  
  public file {  
    name "receptor_pdb"  
    filename "receptor.pdb"  
    path "/"  
    package  
  }  
  cmdline { |ctx| "{0} -l receptor.pdb" }  
}  
  
outputs {  
  public file {  
    name "receptor_pdbqt"  
    filename "receptor.pdbqt"  
    path "/"  
    expected  
  }  
}  
  
prepare_package
```

В листинге 3.3.1, описаны все входные параметры (inputs) и выходные (outputs) со следующими атрибутами:

1. «открытый» (public) – данный атрибут отражает возможность изменения пользователем этого параметра при запуске по средствам скрипта;
2. «ожидаемый» (expected) – данный атрибут отражает, что описанный файл будет рассматриваться системой как результат работы пакета.

3.3.2 Платформенное описание прикладного пакета Autodock Vina как модуля композитного приложения

Для интеграции пакета Autodock Vina в МИТП необходимо описать его взаимодействие с платформой, опираясь на формат представления входных и выходных данных. Платформенный скрипт описания пакета Autodock Vina позволяет определить уровень абстракции и интерпретации в работе с входными и выходными параметрами, что обеспечивает гибкость и упрощение процедуры взаимодействия пользователя с платформой на этапе запуска задания. Тем самым необходимость в определении на входе тех или иных параметров по умолчанию устраняется, а сложные составные входные данные декомпозируются на более мелкие составляющие. Это приводит к более эффективному использованию пакета в рамках состава компонентов выполняемого сервиса.

Платформенный скрипт описания пакета Autodock Vina приведен в листинге 3.3.2.1.

Листинг 3.3.2.1. Основная часть параметров описания пакета Autodock Vina

```

name "vina"
#display_as "vina"
#vendor "Iowa State University"
#url "http://onlinelibrary.wiley.com/doi/10.1002/jcc.21334/abstract"
#license "GPLv3"
#description "Very useful and important package"

inputs {

  public param {
    name "search_space_size_x"
    type float
    default 30.0
  }

  public param {
    name "search_space_size_y"
    type float
    default 30.0
  }

  public param {
    name "search_space_size_z"
    type float
    default 30.0
  }

  public param {
    name "search_space_center_x"
    type float
    default 0.0
  }

  public param {
    name "search_space_center_y"
    type float
    default 0.0
  }

  public param {
    name "search_space_center_z"
    type float
    default 0.0
  }

  public file {
    name "receptor"
    required
    filename "receptor.pdbqt"
    path "/"
    package
  }

  public file {
    name "ligand"
    required
    filename "ligand.pdbqt"
    path "/"
    package
  }

  file {
    name "conf"
    required
    filename "conf.txt"
    path "/"
    depends
    ["search_space_size_x", "search_space_size_y", "search_space_size_z", "search_space_center_x", "search_space_center_y", "search_space_center_z"]
    assembler erb_template rtext("conf.erb")
    package
  }

  cmdline { |ctx| "{0} --config conf.txt --log log.txt" }
}

```

```

}
outputs {
  public file {
    name "ligand_out"
    expected
    filename "ligand_out.pdbqt"
    path "/"
  }

  # public filegroup {
  #   # name "ligand_out.pdbqt"
  #   # filter ["\_out.pdbqt$"]
  #   # path "/"
  # }
}
prepare_package

```

Как видно из листинга 3.3.2.1, все входные параметры описаны со следующими дополнительными атрибутами:

1. «открытый» (public) – данный атрибут отражает возможность изменения пользователем этого параметра при запуске по средствам скрипта;
2. «необходимый» (required) – данный атрибут отражает условие присутствия этого параметра при запуске;
3. «значение по умолчанию» (default) – определяет значение задаваемое по умолчанию, что в данном случае означает задание области поиска с центром в точке (0.0,0.0,0.0). Если использовать значения заданные по умолчанию, докинг будет производиться в области кубической формы с длиной ребра 30 нм.;
4. «тип» (type) – ключевое слово, задающее тип данных, который необходим для отсеивания низкоуровневых ошибок, связанных с типизацией данных.

Листинг 3.3.2.2. Шаблон для конфигурационного файла Autodock Vina

```

receptor = receptor.pdbqt
ligand = ligand.pdbqt

center_x = <%=w.search_space_center_x%>
center_y = <%=w.search_space_center_y%>
center_z = <%=w.search_space_center_z%>

size_x = <%=w.search_space_size_x%>
size_y = <%=w.search_space_size_y%>
size_z = <%=w.search_space_size_z%>

```

Помимо представленных параметров на примере вышеприведенного листинга определяются общие характеристики пакета Autodock Vina.

3.3.3 Платформенное описание прикладного пакета Q region selector

Для интеграции пакета Q region selector в МИТП необходимо описать его взаимодействие с платформой, опираясь на формат представления входных и выходных данных. Платформенный скрипт описания пакета Q_region_selector позволяет определить уровень абстракции и интерпретации в работе с входными и выходными параметрами, что обеспечивает гибкость и упрощение процедуры взаимодействия пользователя с платформой на этапе запуска задания. Тем самым необходимость в определении на входе тех или иных параметров по умолчанию устраняется, а сложные составные входные данные декомпозируются на более мелкие составляющие. Это приводит к более эффективному использованию пакета в рамках состава компонентов выполняемого сервиса.

Платформенный скрипт описания пакета Q_region_selector определяется в соответствии с листингом 3.3.3.

Листинг 3.3.3. Основная часть параметров описания пакета Q_region_selector

```
name "q_region_selector"
display_as "q_region_selector"
vendor "q_region_selector"
license "GPLV3"
description "Selects quantum protein resid near ligand positions"
inputs {
    public file{
        name "ligand_confs"
        required
        package
        filename "ligand.pdbqt"
        path "/"
    }
    public file{
        name "protein"
        required
        package
        filename "protein.pdb"
        path "/"
    }
    cmdline { |ctx| "{0} #{ctx.ligand_confs} #{ctx.protein}" }
}
outputs {
    public file_group {
        name "models"
        filters ["\*.pdb$"]
    }
}
prepare_package
```

Как видно из листинга 3.3.3, все входные параметры описаны со следующими дополнительными атрибутами:

1. «открытый» (public) – данный атрибут отражает возможность изменения пользователем этого параметра при запуске посредством скрипта;

2. «необходимый» (required) – данный атрибут отражает условие присутствия этого параметра при запуске;
3. «пакетный» (package) – данный атрибут отражает условие необходимости передачи данного файла в качестве одного из входных параметров пакета.

3.3.4 Платформенное описание прикладного пакета NWChem как модуля композитного приложения

Для интеграции пакета NWChem в МИТП необходимо описать его взаимодействия с платформой, опираясь на формат представления входных и выходных данных. Платформенный скрипт описания пакета NWChem позволяет определить уровень абстракции и интерпретации в работе с входными и выходными параметрами, что обеспечивает гибкость и упрощение процедуры взаимодействия пользователя с платформой на этапе запуска задания. Тем самым необходимость в определении на входе тех или иных параметров по умолчанию устраняется, а сложные составные входные данные декомпозируются на более мелкие составляющие. Это приводит к более эффективному использованию пакета в рамках выполняемого сервиса.

Платформенный скрипт описания пакета NWChem определяется в соответствии с листингом 3.3.4.

Листинг 3.3.4. Скрипт описания основных параметров пакета NWChem

```
name "nwchem"
#display_as "NWChem"
#vendor "NWCHEM"
#url "http://www.nwchem-sw.org/"
#license "GPLv3"
#description "High-Performance Computational Chemistry to Science"
#logo "/logo.gif"

inputs {
  public file {
    name "config_input"
    required
    package
    filename "input.nw"
    path "/"
  }

  public file {
    name "coordinates_input"
    package
    filename "input.pdb"
    path "/"
  }

  public file {
    name "fragmet_input"
    package
    filename "ZMR.frg"
    path "/"
  }
}
```

```
    cmdline { |ctx| "{0} input.nw -> summary.out" }
}
outputs {
  public file {
    name "summary"
    #display_as "Summary"
    path "/"
    filename "summary.out"
  }
#  public file {
#    name "output_ref"
#    #display_as "Atoms xyz"
#    path "/"
#    filename "output.pdb"
#  }
  public file {
    name "STD_ERR"
    #display_as "STD_ERR"
    path "/"
    filename "std.err"
  }
  public file {
    name "STD_out"
    #display_as "STD_out"
    path "/"
    filename "std.out"
  }
}
prepare_package
```

Как видно из листинга 3.3.4, входные параметры описаны со следующими дополнительными атрибутами:

1. «необходимый» (required) – данный атрибут отражает условие необходимости наличия параметра при запуске задания. В описании пакета NWChem такими параметрами являются максимальное количество итераций (max_iter), молекула заданная в xyz (molecule_xyz) и NWChem_in, являющимся основным файлом конфигурации для пакета NWChem;
2. «тип» (type) – необходим для отсеивания низкоуровневых ошибок, связанных с типизацией данных;
3. «открытый» (public) – данный атрибут отражает возможность изменения пользователем этого параметра при запуске посредством скрипта;
4. «пакетный» (package) – данный атрибут отражает условие необходимости передачи данного файла в качестве одного из входных параметров пакета.

4. ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА

ПС функционирует в рамках распределенной среды облачных вычислений под управлением многофункциональной инструментально-технологической платформы (МИТП) CLAVIRE RU.СНАБ.80066-06. Для использования ПС необходима рабочая станция с подключением к Интернет, со следующими минимальными характеристиками:

- архитектура процессора – x86, x86_64, IA64;
- объем оперативной памяти – 1 ГБ;
- объем свободного пространства на жестком диске – 1 ГБ;
- тактовая частота процессора – 1 ГГц.

Для работы с композитным приложением необходимо использовать браузеры Mozilla FireFox (версия 3.0 и выше), Google Chrome (версия 13 и выше), Opera (версия 9.0 и выше) и Internet Explorer (версия 7.0 и выше).

5. ВЫЗОВ И ЗАГРУЗКА

Исполнение ПС выполняется средствами МИТП CLAVIRE. При запуске композитное приложение описывает следующий процесс, реализуемый в среде облачных вычислений средствами МИТП:

1. По заданному перечню параметров запуска (в тексте программы на языке EasyFlow) определяются входные данные пакетов.
2. На основе мониторинга доступных вычислительных ресурсов VLUC строит оптимальное (с точки зрения минимизации общего времени выполнения) расписание исполнения цепочки запусков.
3. Экземпляр пакета для подготовки структуры рецептора MGL Tools запускается на исполнение на выбранном вычислительном ресурсе.
4. По окончании расчета экземпляра пакета MGL Tools выходные данные (в форме файла) передаются на вход пакету для выполнение задачи докинга Autodock Vina, несколько экземпляров которого параллельно запускаются на исполнение на подготовленных вычислительных ресурсах.
5. По окончании работы всех экземпляров пакета Autodock Vina выходные данные передаются пакету обработки данных Q region selector.

6. После завершения работы пакета Q region selector выходные файлы передаются пакету NWChem.
7. После окончания работы приложения выходные данные передаются в хранилище данных, и пользователь уведомляется средствами МИТП об успешном выполнении задания.

6. ВХОДНЫЕ ДАННЫЕ

Входные данные композитного приложения определяются следующим списком параметров его запуска (см. табл. 6.1). Необходимые промежуточные данные для всех компонентов композитного приложения вычисляются автоматически в процессе его исполнения. Полный набор входных данных состоит из файла содержащего структуру рецептора-мишени и задания области поиска молекулярного сайта.

Таблица 6.1

Входные параметры для композитного приложения

Имя параметра	Описание параметра
Receptor	pdb-файл, содержащий структуру рецептора
search_space_center_x	Центр области поиска молекулярного сайта по координате x
search_space_center_y	Центр области поиска молекулярного сайта по координате y
search_space_center_z	Центр области поиска молекулярного сайта по координате z
search_space_size_x	Длина ребра параллелепипеда, соответствующего области поиска по координате x
search_space_size_y	Длина ребра параллелепипеда, соответствующего области поиска по координате y
search_space_size_z	Длина ребра параллелепипеда, соответствующего области поиска по координате z

7. ВЫХОДНЫЕ ДАННЫЕ

На начальном этапе при помощи пакета MGL Tools заданная пользователем структура подготавливается для процесса докинга и конвертируется в pdbqt формат. Соответствующий файл является результатом работы начального этапа. На следующем этапе отработки пакета Autodock Vina получается столько выходных файлов в формате pdbqt, сколько экземпляров пакета было запущено. Данные передаются пакету Q region

selector, который отбирает структуры по аффинитету, и на выходе отдает выбранные структуры в комплексе со структурой рецептора. Финальный модуль композитного приложения рассчитывает структуру комплекса высокоточными методами и предоставляет результат в отдельном файле, который передается пользователю после завершения работы приложения.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

ПС	Прикладной сервис
МИТП	Многофункциональная инструментально-технологическая платформа

ПЕРЕЧЕНЬ ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. [Электронный ресурс]: <<http://mgltools.scripps.edu/documentation>>.
2. [O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry* 31 \(2010\) 455-461](#)
3. [Электронный ресурс]: <<http://www.thch.uni-bonn.de/tc/NWChem/>>.

