

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

СОГЛАСОВАНО
Генеральный директор
ЗАО «АйТи»



УТВЕРЖДАЮ
Ректор НИУ ИТМО



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

ПРИКЛАДНОЙ СЕРВИС ВЫЧИСЛЕНИЯ ПОКАЗАТЕЛЯ КИСЛОТНОСТИ
СРЕДСТВАМИ КВАНТОВОЙ ХИМИИ

ОПИСАНИЕ ПРОГРАММЫ

ЛИСТ УТВЕРЖДЕНИЯ

RU.СНАБ.80066-06 13 53-ЛУ

Инев. № годл.	Подп. и дата
Взам. инв. №	Инев. № дубл.
Подп. и дата	Подп. и дата

Представители
Организации-разработчика

Руководитель разработки,
профессор НИУ ИТМО

Бухановский А.В.

“20” декабря 2011 г.

Ответственный исполнитель,
с.н.с. НИУ ИТМО

Луценко А.Е.

“20” декабря 2011 г.

Нормоконтролер
ведущий инженер НИУ ИТМО

Позднякова Л.Г.

“20” декабря 2011 г.

2011

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**УТВЕРЖДЕН
RU.СНАБ.80066-06 13 53-ЛУ**

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

**ПРИКЛАДНОЙ СЕРВИС ВЫЧИСЛЕНИЯ ПОКАЗАТЕЛЯ КИСЛОТНОСТИ
СРЕДСТВАМИ КВАНТОВОЙ ХИМИИ**

ОПИСАНИЕ ПРОГРАММЫ

RU.СНАБ.80066-06 13 53

ЛИСТОВ 28

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

2011

АННОТАЦИЯ

Документ содержит описание прикладного сервиса вычисления показателя кислотности средствами квантовой химии. Прикладной сервис реализует композитное приложение для многоцелевой инструментально-технологической платформы (МИТП) CLAVIRE, которое позволяет, основываясь на заданной структуре атомно-молекулярной системы, рассчитать показатель кислотности рКа. 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	4
2.1.	Область применения	4
2.2.	Функциональное назначение.....	5
2.3.	Ограничения на применение	5
3.	ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ	5
3.1.	Структура композитного приложения	5
3.2.	Характеристики модулей композитного приложения	8
3.3.	Описание модулей композитного приложения в МИТП	12
4.	ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА	19
5.	ВЫЗОВ И ЗАГРУЗКА.....	19
6.	ВХОДНЫЕ ДАННЫЕ	22
7.	ВЫХОДНЫЕ ДАННЫЕ	24
	ПЕРЕЧЕНЬ СОКРАЩЕНИЙ	26
	ПЕРЕЧЕНЬ ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	27

1. ОБЩИЕ СВЕДЕНИЯ

Прикладной сервис (ПС) вычисления показателя кислотности средствами квантовой химии RU.СНАБ.80066-06 13 53 реализует композитное приложение, которое позволяет, основываясь на заданной структуре атомно-молекулярной системы, рассчитать показатель кислотности pK_a (т.е. взятое с отрицательным знаком значение логарифма константы диссоциации). 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):

ORCA - прикладной пакет, позволяющий проводить высокоточные квантово-химические расчеты (заимствуется);

GAMESS - прикладной пакет, реализующий разнообразные квантово-химические методы (заимствуется);

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

2. ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ

2.1. Область применения

Задача предсказания констант диссоциации, прежде всего, позволяет прогнозировать устойчивость наноструктур в водных средах, поскольку способность нанообъектов растворяться в воде и устойчивость их растворов определяются наличием, в частности, кислотных диссоциирующих групп на их поверхности. С другой стороны, если

эти группы находятся в состоянии диссоциации, то имеющийся электрический заряд оказывает существенное влияние на оптические и электрические свойства нанобъектов. Поэтому знание состояния диссоциации данных групп является одним из первостепенных вопросов при моделировании нанобъектов в водных средах. Кроме того, изменяя рН среды и, соответственно, изменяя состояние диссоциации указанных групп, можно направленным образом изменять оптические, фотофизические и электрические свойства нанобъектов, что также может быть предметом моделирования молекулярно-динамическими или квантово-химическими методами.

2.2. Функциональное назначение

Сервис предназначен для теоретического расчета величины показателя кислотности (pK_a) квантово-химическими методами *ab initio* для произвольной атомно-молекулярной системы. Сервис применим к системам с любым количеством, типом и характером диссоциирующих групп: величина pK_a определяется для любой заданной группы.

2.3. Ограничения на применение

В силу того, что сервис основан на применении квантово-химических методов *ab initio*, его использование для систем, содержащих более ~100 атомов является принципиально возможным, но может оказаться затруднительным из-за слишком длительного времени счета.

3. ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ

3.1. Структура композитного приложения

Работа сервиса включает в себя в два этапа. На первом этапе рассчитывается полная энергия для исходного состояния молекулярной системы. Эта задача, в свою очередь, решается в две последовательные стадии. На 1-й стадии решается задача оптимизации геометрии молекулы методом DFT/B3LYP в базисе 6-31G(d). Для этого применяется пакет ORCA. На 2-й стадии проводится расчет полной энергии для оптимизированной геометрии тем же методом, но в более широком базисе: 6-31++G(2d,p) и с учетом сольватации (модель PCM с параметрами для воды в качестве растворителя). Для этого применяется пакет GAMESS.

На втором этапе определяется полная энергия для аниона исследуемой молекулярной системы, получающегося в результате диссоциации той группы, для которой рассчитывается pK_a . Расчет проводится также в две стадии и теми же методами и с применением тех же пакетов, которые использовались на первом этапе.

Результат – значение pK_a – может быть получен по формуле: $pK_a = -6.9 + 0.55 \cdot (pK_a)_{расч.}$, где расчетное значение $(pK_a)_{расч.}$ определяется формулой

$$(pK_a)_{расч.} = \frac{1}{2.3RT} (D - (5/2)RT - 259.5 \text{ ккал/моль}),$$

D – разность полных энергий, полученных на аниона и молекулы, R – универсальная газовая постоянная, T – температура в °К.

Принципиально важно, что первый и второй этапы являются независимыми друг от друга и могут выполняться параллельно. Учитывая это, композитное приложение обеспечивает следующий процесс, реализуемый средствами МИТП-Ц.

По заданному перечню параметров запуска (в тексте программы на языке EasyFlow) определяются входные данные пакетов ORCA и GAMESS для двух параллельных веток: нейтральной молекулярной системы и ее аниона.

На основе мониторинга доступных вычислительных ресурсов в корпоративной среде МИТП-Ц строит оптимальное (с точки зрения минимизации общего времени выполнения) расписание исполнения цепочек запусков.

Экземпляры пакета ORCA для двух систем с уточнением параметров метода, применяемого при оптимизации геометрии, запускаются на исполнение на выбранных вычислительных ресурсах; в процессе исполнения осуществляется мониторинг их состояния.

По окончании расчета экземпляров пакета ORCA выходные данные расчета оптимизации геометрии обеих систем передаются на вход пакету GAMESS, экземпляры которого запускаются на исполнение на подготовленных вычислительных ресурсах; в процессе исполнения осуществляется мониторинг их состояния выполнения.

По окончании расчета выходные данные передаются в хранилище данных МИТП-Ц и пользователь уведомляется средствами МИТП-Ц об успешном выполнении задания.

На листинге 1 приведен текст композитного приложения на языке EasyFlow, соответствующий перечисленным выше действиям.

Листинг 1. Текст композитного приложения на языке EasyFlow

```
require xyzNormalMolecule;
require xyzAnionMolecule;
step NormalGeom runs ORCA
```

```

(
    calc_method = "DFT",
    basis = "6-31G(d)",
    functional = "B3LYP",
    task = "geometry optimization",
    molecule_xyz = xyzNormalMolecule
);
step NormalEnergy runs Gamess
(
    calc_method = "DFT",
    solvent = "water",
    basis = "6-311++G(2d,p)",
    functional = "B3LYP",
    task = "single point",
    molecule_xyz = NormalGeom.Result.outs["orca_xyz.out"]
);
step AnionGeom runs ORCA
(
    calc_method = "DFT",
    charge = -1,
    basis = "6-31G(d)",
    functional = "B3LYP",
    task = "geometry optimization",
    molecule_xyz = xyzAnionMolecule
);
step AnionEnergy runs Gamess
(
    calc_method = "DFT",
    solvent = "water",
    basis = "6-311++G(2d,p)",
    charge = -1,
    functional = "B3LYP",
    task = "single point",
    molecule_xyz = AnionGeom.Result.outs["orca_xyz.out"]
);

```

На указанном листинге блок “step NormalGeom runs ORCA.DFT” обеспечивает расчет оптимизированной геометрии для нейтральной молекулы методом DFT/B3LYP/6-31G*; блок “step NormalEnergy runs Gamess.DFT” обеспечивает расчет полной энергии для оптимизированной геометрии нейтральной молекулы методом DFT/B3LYP/6-31++G(2d,p). Блоки “step NormalAnion runs ORCA.DFT” и “step NormalAnion runs Gamess.DFT” обеспечивают то же самое для аниона. Из приведенного на листинге 1 текста также видно, что данное приложение требует для своей работы два входных файла (см. первые две строчки листинга): файл координат атомов нейтральной молекулы (require xyzNormalMolecule) и файл координат атомов аниона (require xyzAnionMolecule). Оба файла должны быть представлены в xyz-формате, примеры таких файлов приведены ниже (см. раздел 8).

На рис.1 результат интерпретации приведенного листинге 1 скрипта в МИТП, который создает две параллельные ветки из двух последовательных процессов запуска пакетов каждая.

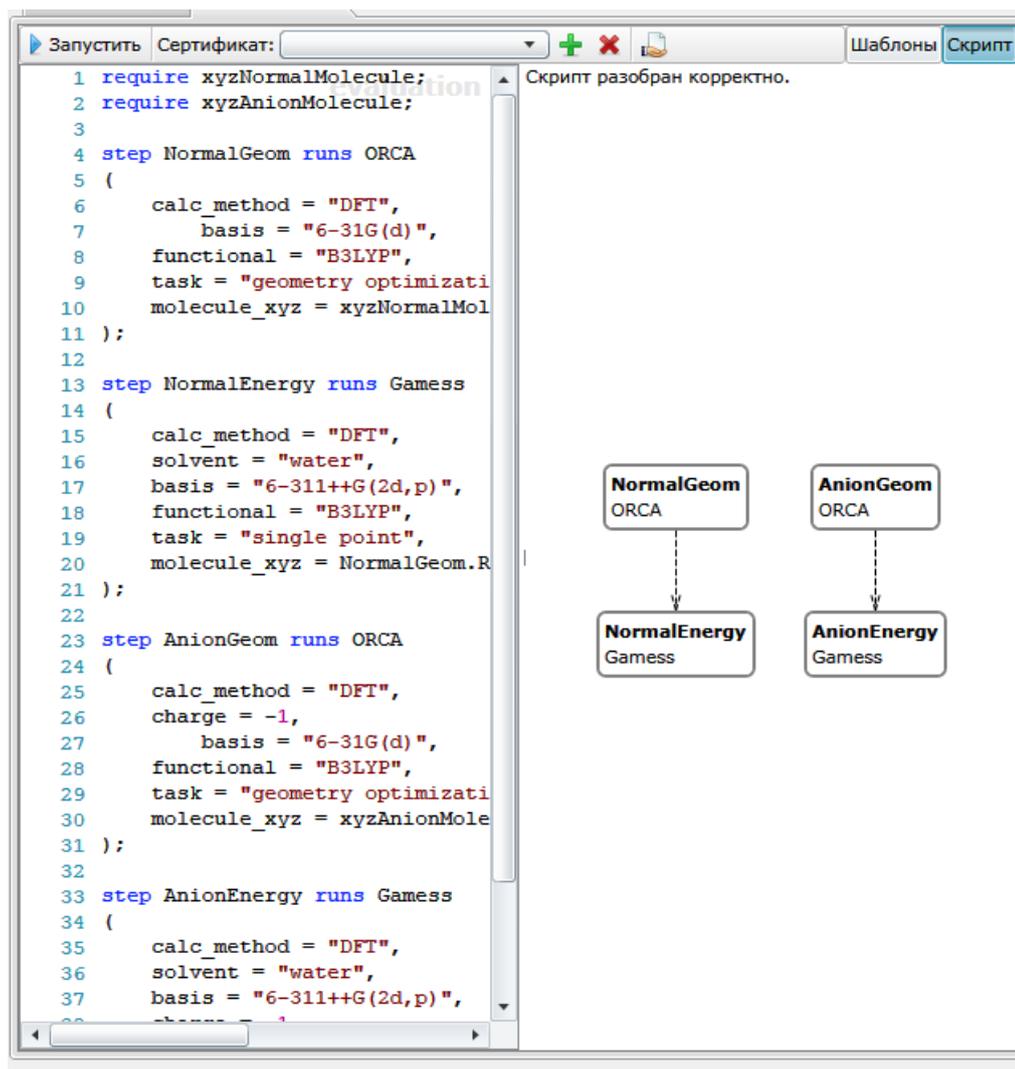


Рисунок 1 - Результат интерпретации скрипта, приведенного листинге 1

3.2. Характеристики модулей композитного приложения

3.2.1. Прикладной пакет квантовой химии ORCA

Прикладной пакет ORCA является программным компонентом, позволяющим производить высокоточные расчеты атомно-молекулярных систем, рассчитывать характеристики как основного, так и возбужденных состояний. Пакет ORCA универсальным квантово-химическим пакетом, реализующим набор высокоточных методов квантовой химии, позволяющих моделировать основное и возбужденные состояния произвольных молекулярных систем. В числе прочих, реализуется набор методов, учитывающих эффекты межэлектронной корреляции, высокоэффективные методы, позволяющие оптимизировать геометрию, рассчитывать колебательные (инфракрасные и рамановские) спектры. Более подробное описание возможности пакета и его структуры можно найти в документации пакета [1].

Основные типовые задачи, решаемые пакетом, следующие:

- (1) Задачи «одной точки» (т.е. фиксированной геометрии) для основного состояния, в т.ч. задачи вычисления полной энергии разнообразными высокоточными методами с учетом электронных корреляций.
- (2) Задачи типа оптимизации геометрии, в т.ч. задачи определения равновесной геометрии, поиска седловой точки (т.е. определения высоты потенциального барьера) и вычисления колебательных частот молекул.
- (3) Задачи расчета энергий электронно-возбужденных состояний и сил осцилляторов электронных переходов.

Пакет реализует методы, базирующиеся на методах Хартри–Фока, функционала плотности и многоконfigurационного самосогласованного поля. Данный программный компонент позволяет моделировать структуры как с замкнутой, так и с открытой оболочкой.

Пакет ORCA является внешней разработкой, реализованной на языке C/C++. Дистрибутивы пакета существуют как для семейства операционных систем Windows, так и для Linux.

Пакет ORCA 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.

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

```
orca Входной_файл.inp > Выходной_файл.out.
```

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

Входным параметром пакета является входной файл задачи. Пример входного файла приведен на листинге 2. Описание расчетной задачи включает в себя указание метода расчета и его параметров, а также задание количества параллельных узлов. Данное описание занимает несколько первых строк входного файла, всегда предваряется восклицательным знаком «!» и осуществляется при помощи ключевых слов,

описывающих основные параметры запуска пакета и молекулярную систему в трехмерном пространстве. После описания расчетной задачи следует описание рассчитываемой молекулярной системы. Описание системы представляет собой массив из четырех столбцов, каждая строка которого описывает один из атомов молекулы (xyz-формат). В первом столбце задается тип атома, во втором, третьем и четвертом его координаты по x, y и z соответственно. Количество электронов в атоме жестко связано с тем, к какому типу атомов он принадлежит, а значит, определяется заданием первого столбца. Массив, описывающий атомы в молекуле, предваряется инструкцией «* xyz 0 1» (см. листинг 2). Более подробно это изложено в пользовательской инструкции [2].

Листинг 2. Пример входного файла пакета ORCA

```
! RKS B3LYP 6-31G TightSCF Opt NumFreq
%geom MaxIter 600
end
* xyz 0 1
O      4.0      5.70774      -7.08272
H      3.0      6.0      -7.74249
H      4.0      5.0      -7.51330
*
```

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.

3.2.2. Прикладной пакет квантовой химии GAMESS

Прикладной пакет GAMESS представляет собой открытый проект, реализующий большое число квантово-химических методов и подходов. Одной из его особенностей являются широкие возможности учета влияния растворителя в квантовой химии. В частности, он позволяет моделировать эффекты растворителя с использованием модели поляризуемого континуума, что непосредственно используется при решении задачи предсказания константы депротонирования.

Данный пакет решает принципиально те же типовые задачи, что и пакет ORCA (см. раздел 3.2.1), в частности, задачи «одной точки», задачи типа оптимизации геометрии, задачи расчета характеристик основного и возбужденных состояний.

Пакет GAMESS способен производить расчеты на базе набора высокоточных методов квантовой химии, позволяющих моделировать основное и возбужденные состояния молекулярных систем, включая:

- метод Хартри–Фока, теорию функционала плотности,
- многоконfigurационное самосогласованное поле,
- обобщенный метод валентных связей,
- учет электронной корреляции по методу Меллера–Плессета второго порядка,
- конфигурационное взаимодействие, позволяющее провести коррекцию энергии электронной корреляции после процедуры самосогласования,
- учет различных поправок и эффектов.

Для более детального ознакомления с функциональностью, структурой и особенностями применения пакета можно воспользоваться документацией [3].

Пакет GAMESS является внешней разработкой, реализованной на языке Fortran. Дистрибутивы пакета существуют как для семейства операционных систем Windows, так и для Linux.

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 POSIX-совместимых операционных системах, OS X и Windows, также ряд других платформ Sun Solaris UltraSPARC/x86/x86-64, Linux x86/x86-64 и т.д.

Основные программные модули GAMESS поддерживают параллельный режим счета как на многопроцессорных компьютерах, так и в кластерах рабочих станций UNIX. Основой параллелизации кода является язык послышки сообщений.

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

```
runqms Входной_файл_без_расширения > Выходной_файл.out.
```

В случае удачного запуска и отработки приложения в директории появится файл Выходной_файл.out. с ВЫХОДНЫМИ ДАННЫМИ.

Входным параметром пакета является входной файл задания. Пример такого файла приведен на листинге 3. Описание расчетной задачи обычно включает в себя описание основных химических параметров, параметров настройки ресурсов исполнения, параметров базиса. После описания расчетной задачи следует описание рассчитываемой молекулярной системы. Описание системы определяется следующими возможными основными типами координат атомов: декартовы координаты, координаты Хильдербрандта, Z-матрицы в стандарте MOPAC и GAUSSIAN, а также собственный способ задания Z-матрицы. Декартовое описание представляет собой массив из пяти столбцов, каждая строка которого описывает один из атомов молекулы. В первом столбце задается тип атома, во втором номер в периодической таблице, далее – координаты по x, y и z (см. листинг 3). Более подробно это изложено в пользовательской инструкции [4].

Листинг 3. Пример входного файла для пакета GAMESS

```
! File created by the GAMESS Input Deck Generator Plugin for Avogadro
$CONTRL SCFTYP=RHF RUNTYP=ENERGY DFTTYP=B3LYP MAXIT=100 ICHARG=-1 $END
$BASIS GBASIS=N311 NGAUSS=6 NDFUNC=2 NPFUNC=1 DIFFSP=.TRUE.
      DIFFS=.TRUE. $END
$PCMC SOLVNT=WATER $END
$DATA
Title
C1
C      6.0  -0.8767411777   0.8676677337   0.0000000000
O      8.0   0.1736489859   1.5531019130   0.0000000000
O      8.0  -1.0518588905  -0.3742749433   0.0000000000
H      1.0  -1.8494789177   1.4866952965   0.0000000000
$END
```

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.

3.3. Описание модулей композитного приложения в МИТП

3.3.1. Прикладной пакет квантовой химии ORCA

Для интеграции пакета ORCA в МИТП необходимо описать на языке EasyPackage его взаимодействие с платформой, опираясь на формат представления входных и выходных данных. Платформенный скрипт описания пакета ORCA позволяет определить уровень абстракции и интерпретации в работе с входными/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. Еще одним важным фактором использования платформенного скрипта описания пакета ORCA является возможность извлечения только необходимой информации из выходных данных для дальнейшей обработки как пользователем, так и последующим выполняемым блоком сервиса.

Платформенный скрипт описания пакета ORCA определяется в соответствии с листингом 4.

Листинг 4. Скрипт описания основных параметров пакета ORCA.

```
meta param {
  name "max_iters"
  required
  display_as "Maximum number of iterations for geometry optimization"
  type int
  default 600
  validator lambda { |val, ctx| val > 0 }
}
meta param {
  name "method"
  ontology "escience.ifmo.ru/onto/method"
  display_as "Method"
  type enum ["HF", "DFT"]
  default "DFT"
}
meta param {
  name "functional"
  ontology "escience.ifmo.ru/onto/functional"
  display_as "Functional"
  type enum ["B3LYP"]
  default "B3LYP"
}
meta param {
  name "task"
  ontology "escience.ifmo.ru/onto/task"
  display_as "Task"
  type enum ["optimization", "single point"]
  default "optimization"
}
meta param {
  name "basis"
  display_as "Basis"
  type string #enum ["6-31G", "6-31G*", "6-31G**", "6-31++G**"]
  default "6-31G"
  depends_on [":basis"]
}
auto param {
  name "system_charge"
  type int
  default 0
}
meta file {
  name "molecule_xyz"
  required
  display_as "Molecule in XYZ format"
  extractor XYZFileExtractor.new
}
meta param {
  name "atoms_xyz"
  display_as "Atoms XYZ"
```

```

    type string
    default ""
  }
  raw file {
    name "orca_in"
    required
    place "/"
    filename "orca.in"
    assembler erb_template rtext("orcaFullEnergy.erb")
  }
  cmdline lambda { |ctx| "{0} orca.in orca.out" }
}
outputs {
  meta param {
    name "final_energy_eval_xyz"
    display_as "Atoms XYZ"
    type string
    default ""
  }
  meta file {
    name "orca_xyz_out"
    filename "orca_xyz.out"
    place "/"
    assembler ObjectToSAssembler.new("final_energy_eval_xyz")
  }
  raw file {
    name "orca_out"
    filename "orca.out"
    place "/"
    extractor XYZOrcaResultExtractor.new
  }
}

```

Как видно из листинга 4, входные параметры описаны со следующими дополнительными атрибутами:

«необходимый» (required) – данный атрибут отражает условие необходимости наличия параметра при запуске задания; в описании пакета ORCA такими параметрами являются максимальное количество итераций (max_iter), молекула заданная в xyz-формате (molecule_xyz) и orca_in, являющийся основным файлом конфигурации для пакета ORCA;

«информация о параметре» (Display_as) – данный атрибут отражает текстовую информацию о параметре в интерфейсе пользователя;

«значение по умолчанию» (default) – определяет начальное значение, в случае с method – установлено DFT, т.е. выбранный для выполнения метод расчета, аналогичные рассуждения с task, и с functional. Начальное значение system_charge по default установлено нулевым, что соответствует нормальной, незаряженной молекулярной системе;

«валидатор» (validator) – необходим для проверки валидности значения параметра, например, у max_itr установленное значение переменной должно быть больше нуля;

«тип» (type) – необходим для отсеивания низкоуровневых ошибок, связанных с типизацией данных;

«экстракторы и ассемблеры» – атрибуты файлов, необходимые для нахождения какой-либо информации (конкретного значения, строки результата) внутри файла и создания новых файлов на основе указанных значений параметров и описанных шаблонов. Экстрактор результата работы ORCA для нахождения хуз молекулы с вычисленными значениями задачи оптимизации описан в листинге 5.

Листинг 5. XYZ-экстрактор из результатов расчета ORCA

```

class XYZOrcaResultExtractor < TextExtractor

  @format = "xyz"

def extract(str)
  xyz = str.to_s
  #xyz.delete! "\r"
  lines = xyz.split("\n")
  mode="search final energy"
  Params = {}
  Params["final_energy_eval_xyz"] = ""

  linesToMiss = 5;
  # FSM
  isFirst = true
  lines.each { |line|
    line.strip!
    if mode == "search final energy"
      foundHeadLine = line.scan("FINAL ENERGY EVALUATION AT THE STATIONARY POINT")
      if (foundHeadLine != [])
        mode = "result block found"
      end
    elsif mode == "result block found"
      linesToMiss -= 1
      if (linesToMiss == 0)
        mode = "energy molecule table"
      end
    elsif mode == "energy molecule table"
      foundSamples = line.scan(/^(w+)\s+([\d\-.]+)\s+([\d\-.]+)\s+([\d\-.]+)\s*$/i)
      if (foundSamples == [])
        mode = "ignore"
      else
        number = "0"
        if (foundSamples[0][0] == "O")
          number = "8"
        elsif (foundSamples[0][0] == "H")
          number = "1"
        end
      end

      if (isFirst)
        isFirst = false
      else
        Params["final_energy_eval_xyz"] += "\n"
      end
      Params["final_energy_eval_xyz"] += foundSamples.at(0).at(0) + " " + number +
      " " + foundSamples.at(0).at(1) + " " + foundSamples.at(0).at(2) + " " +
      foundSamples.at(0).at(3)
    end
  end
}

$log.Trace(Params["final_energy_eval_xyz"].to_s)

puts Params
return Params

```

```
end
end
```

Подобным образом определяются и ассемблеры, задача которых – собрать входной файл по указанному шаблону (см. листинг 6).

Листинг 6. Шаблон для построения входного файла ORCA

```
! RKS <%= w.functional %> <%= w.basis %> TightSCF Opt NumFreq
%geom MaxIter <%= w.max_iters %>
end
* xyz <%= w.system_charge %> 1
<%= w.atoms_xyz %>
*
```

Помимо представленных параметров на примере приведенного выше скрипта определяются общие характеристики пакета ORCA.

3.3.2. Прикладной пакет квантовой химии GAMESS

Для интеграции пакета GAMESS в МИТП необходимо описать его взаимодействия с платформой, опираясь на формат представления входных и выходных данных. Платформенный скрипт описания пакета GAMESS позволяет определить уровень абстракции и интерпретации в работе с входными/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. Еще одним важным фактором использования платформенного скрипта описания пакета GAMESS является возможность извлечения только необходимой информации из выходных данных для дальнейшей обработки как пользователем, так и последующим выполняемым блоком сервиса.

Платформенный скрипт описания пакета GAMESS определяется в соответствии с листингом 7.

Листинг 7. Основная часть параметров описания пакета GAMESS

```
inputs {
  meta param {
    name "method"
    ontology "escience.ifmo.ru/onto/method"
    display_as "Method"
    type enum ["HF", "DFT"]
    default "DFT"
  }
  meta param {
    name "functional"
```

```

ontology "escience.ifmo.ru/onto/functional"
display_as "Functional"
type enum ["B3LYP"]
default "B3LYP"
}
meta param {
name "task"
ontology "escience.ifmo.ru/onto/task"
display_as "Task"
type enum ["optimization", "single point", "energy"]
default "optimization"
}

meta param {
name "basis"
display_as "Gamess basis"
type string #enum ["6-31G", "6-31G*", "6-31G**", "6-31++G**"]
default "GBASIS=N311 NGAUSS=6 NDFUNC=2 NPFUNC=1 DIFFSP=.TRUE. DIFFS=.TRUE."

raw param {
name "runtype"
display_as "Run type"
type enum ["OPT", "SP", "ENERGY"]
default "ENERGY"
evaluator lambda { |ctx| true}
}
}
meta param {
name "system_charge"
type int
default 0
}
}
meta param {
name "multiplicity"
type int
}
}
meta param {
name "iswater"
type bool
default true
}
}
meta param {
name "hssend"
display_as "hssend?"
type bool
default true
}
}
meta file {
name "molecule_xyz"
display_as "Molecule in XYZ"
extractor XYZGamessFileExtractor.new
}
}
meta param {
name "atoms_xyz"
display_as "Atoms XYZ"
type string
default ""
}
}
raw file {
name "gamess_input"
required
filename "gamess.inp"
place "/"
assembler old_school_template rtext("template.ost")
}
}

cmdline lambda { |ctx| "{0} gamess.inp gamess.out" }
}

```

```

outputs {
  meta param {
    name "total_energy"
    type string
    default ""
    validator lambda { |val, ctx| true }
  }
  auto file {
    name "gamess_out"
    filename "gamess.out"
    place "/"
    extractor TotalEnergyGamessExtractor.new
  }

  meta file {
    name "total_energy_out"
    filename "total_energy.out"
    place "/"
    assembler ObjectToSAssembler.new("total_energy")
  }
}

```

Как видно из листинга 7, входные параметры описаны со следующими дополнительными атрибутами:

«информация о параметре» (`Display_as`) – данный атрибут отражает текстовую информацию о параметре в интерфейсе пользователя;

«значение по умолчанию» (`default`) – определяет начальное значение, в случае с `method` – установлено `DFT`, т.е. выбранный для выполнения метод расчета, аналогичные рассуждения с `task`, `functional`, `basis`, `iswater`, `hssend` и др. Начальное значение `system_charge` по `default` установлено в 0, что соответствует нормальной, незаряженной молекулярной системе;

«тип» (`type`) – необходим для отсеивания низкоуровневых ошибок, связанных с типизацией данных;

«экстракторы и ассемблеры» – атрибуты файлов, необходимые для нахождения какой-либо информации (конкретного значения, строки результата) внутри файла и создания новых файлов на основе указанных значений параметров и описанных шаблонов. Экстрактор результата работы GAMESS для нахождения полной энергии молекулы представлен в листинге 8.

Листинг 8. Экстрактор полной энергии из результатов расчета GAMESS

```

class TotalEnergyGamessExtractor < TextExtractor
  @format = "xyz"
  def extract(str)
    xyz = str.to_s
    lines = xyz.split("\n")
    mode="searchTotalEnergy"

    Params = {}
    Params["total_energy"] = ""
    linesToMiss = 8;
    # FSM

```

```
lines.each { |line|
  line.strip!
  if mode == "searchTotalEnergy"
    foundEnergyBlock = line.scan("ENERGY COMPONENTS")
    if (foundEnergyBlock != [])
      mode = "energyBlockFound"
    end
  elsif mode == "energyBlockFound"
    linesToMiss -= 1
    if (linesToMiss == 0)
      mode = "totalEnergyLine"
    end
  elsif mode == "totalEnergyLine"
    foundSample = line.scan(/([\d\-\.\.]+)\s*$/i)
    Params["total_energy"] = foundSample.at(0).at(0)
    mode = "ignore"
  end
}
puts Params
return Params
end
end
```

Помимо представленных параметров на примере вышеприведенного скрипта определяются также общие характеристики пакета GAMESS.

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

Вызов и загрузка композитного приложения выполняются средствами МИТП-Ц CLAVIRE и компонента взаимодействия с пользователем RU.СНАБ.80066-06 01 21. После запуска из интерфейса пользователя и вставления текста скрипта композитного приложения, приведенного на листинге 1, производится внутриплатформенная

интерпретация языка EasyFlow, которая создает две параллельные очереди из двух последовательных процессов запуска пакетов ORCA и GAMESS.

Интерфейс разделен на несколько рабочих секций. Каждая секция озаглавлена в соответствии с ее назначением. Название секции находится над рабочим полем данной секции. В интерфейсе представлены три основные рабочие секции: это секция «проекты», секция «новая задача», которую пользователь может переименовать по своему желанию в соответствии и названием своей задачи и секция «файлы». После загрузки интерфейса комплекса, в секции «проекты», расположенной на интерфейсе слева, будет автоматически создан проект, задача и, возможно, там будут находиться некоторые уже загруженные файлы, пригодные, например, для тестовых расчетов. В центральной части интерфейса находится поле, в котором оказывается введенным скрипт решения данной задачи. В данном случае это текст композитного приложения (см. листинг 1). В правом окне интерфейса, озаглавленного как «Входные данные», появятся ярлыки обозначающие входные данные.

Теперь следует загрузить файлы исходных данных, необходимые для решения данной задачи. Для этого в секции «файлы» в плавающем меню, активизируемом правой клавишей мыши, нужно выбрать «добавить файл» (рис.2).

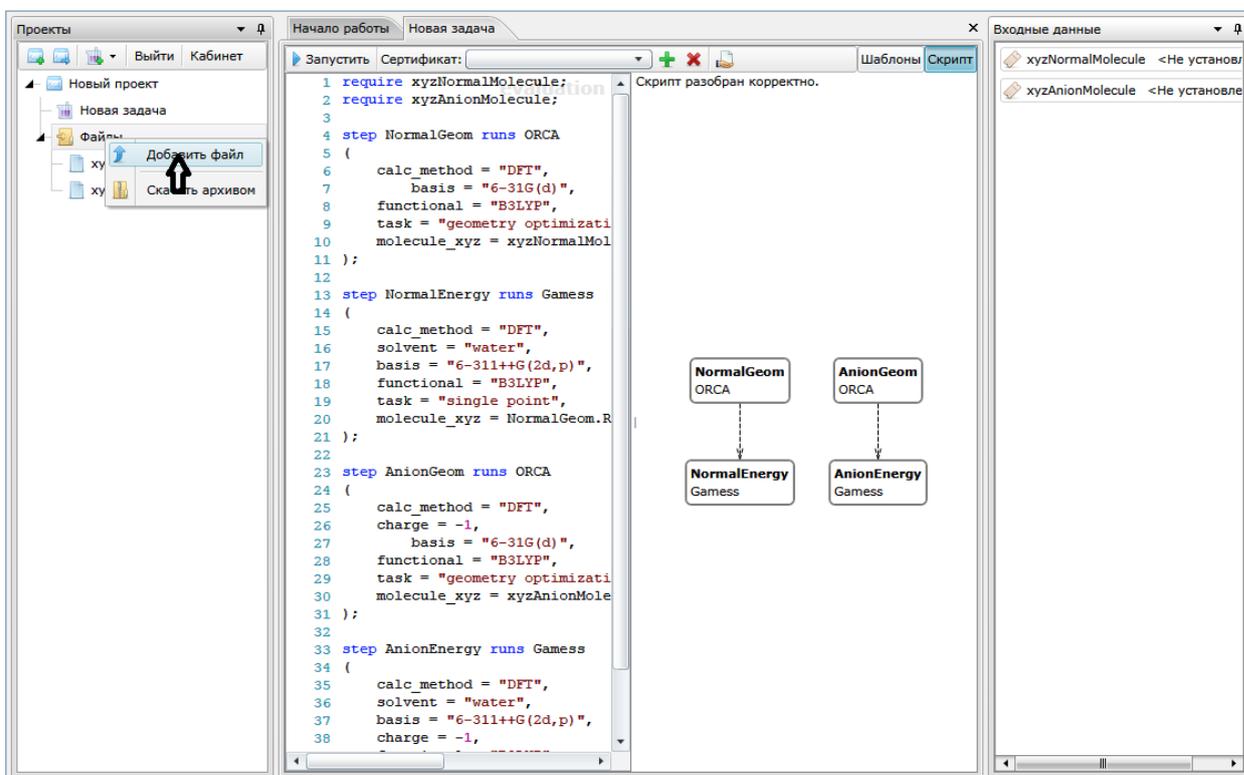


Рисунок 2 - Загрузка файлов исходных данных

После того, как будут загружены оба входных файла координат, упомянутых в разделе 3.1, нужно связать эти файлы с параметрами задачи. Это выполняется в правом

окне (рис.3) путем выбора нужного имени файла в плавающем меню, появляющемся после нажатия правой клавиши мыши на имени параметра. При этом важно, чтобы входной файл координат атомов нейтральной молекулы был связан с параметром «xyzNormalMolecule», а файл координат аниона – с именем «xyzAnionMolecule», а не наоборот.

После выполнения ассоциации входных файлов с параметрами задачи следует запустить задачу, как показано на рис. 4.

О завершении всех четырех этапов решения задачи пользователь извещается, как показано на рисунке 5. При этом в левой части интерфейса появляются ярлыки, соответствующие выходным файлам. Эти выходные файлы следует скачать в какую-либо локальную директорию, для чего следует в плавающем меню, открывающемся на имени файла нажатием правой клавиши мыши, выбрать «скачать».

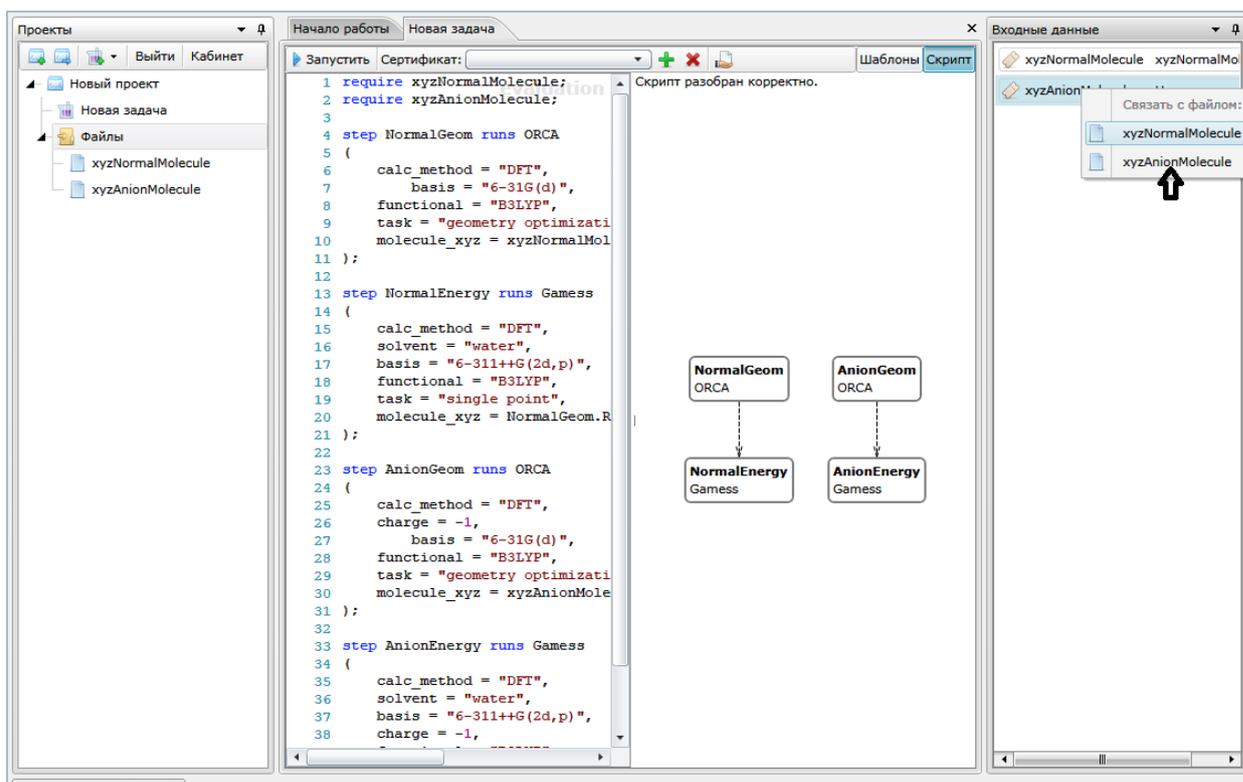


Рисунок 3 - Связывание загруженных файлов с параметрами задачи.

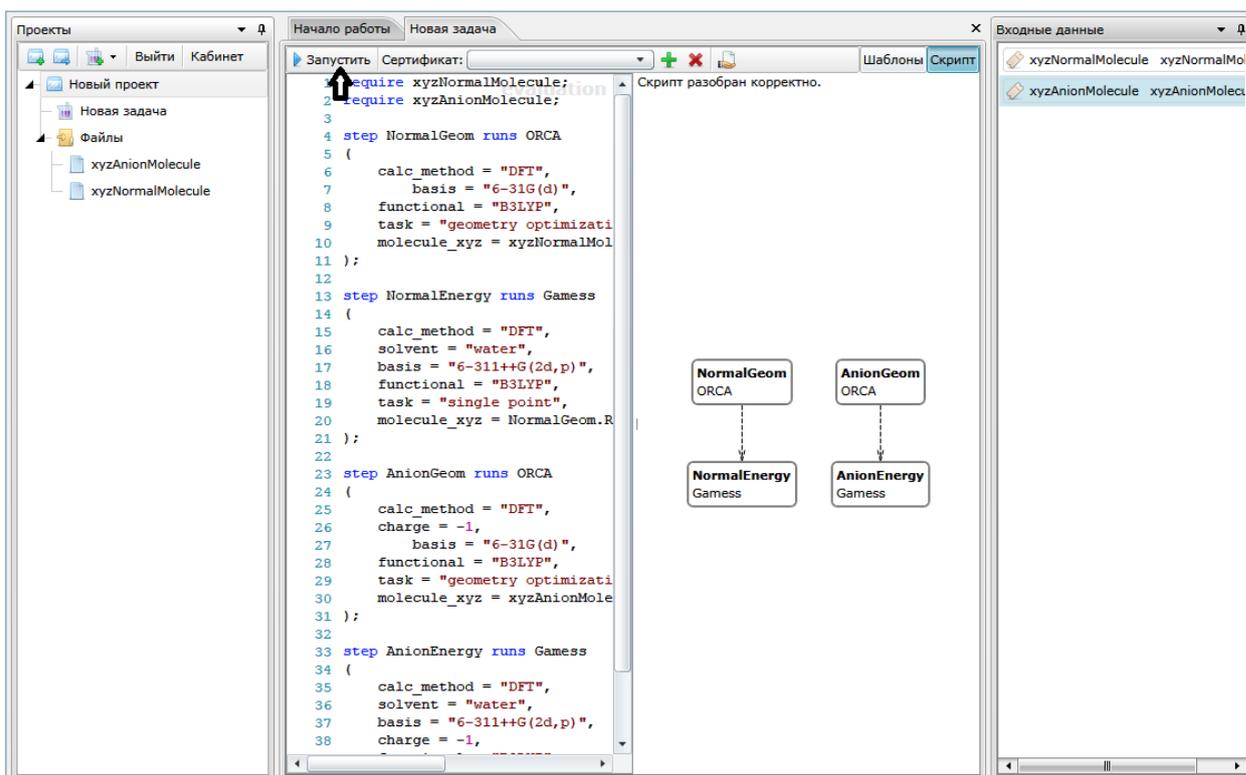


Рисунок 4 - Запуск задачи.

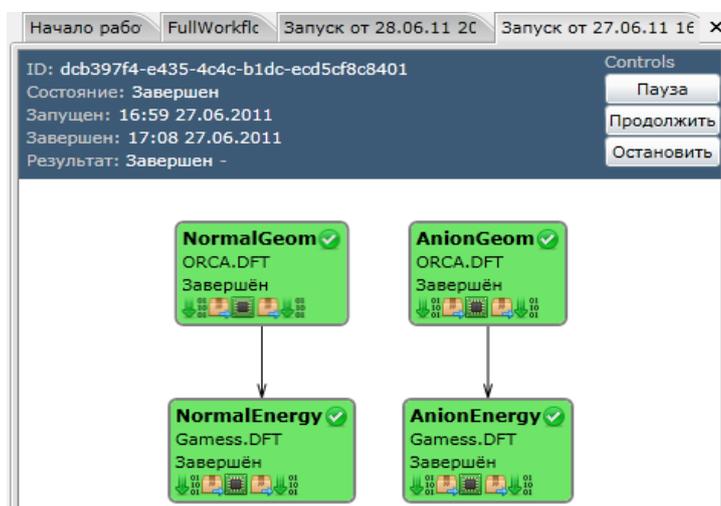


Рисунок 5 - Завершение задачи.

6. ВХОДНЫЕ ДАННЫЕ

Комплект входных данных для данного сервиса включает в себя (1) файл координат атомов для исследуемой молекулы в xyz-формате, (2) файл координат атомов для аниона в xyz-формате и (3) дополнительные параметры (не обязательно).

Файл координат атомов исследуемой молекулы может быть получен с помощью любого доступного молекулярного редактора, например, Avogadro [5]. Пример такого файла для молекулы динитрамида $\text{NH}(\text{NO}_2)_2$ приведен на листинге 9.

Листинг 9. Файл координат атомов для молекулы динитрамида в xyz-формате

```

8
Dinitramide
N      -1.57397      2.01010      -0.06057
N      -0.43868      1.58643       0.49954
N       0.61886      2.23572       0.02190
H      -0.35923      0.67156       0.07907
O      -2.25481      1.09652      -0.52808
O      -1.85105      3.20784      -0.00858
O       1.00030      3.21935       0.65407
O       1.17194      1.67561      -0.92500

```

Число на 1-й строчке – число атомов. 2-я строчка – для комментария. Начиная с 3-й строчки собственно координаты атомов (в ангстремах). Каждая строчка соответствует одному атому. Для каждого атома указывается: химический символ атома, x-координата, y-координата и z-координата.

Файл координат атомов аниона отличается файла для молекулы отсутствием строки, соответствующей диссоциирующему атому водорода, т.е. тому атому водорода, для диссоциации которого и рассчитывается значение рКа. Соответственно, на единицу должно быть уменьшено число атомов. Пример файла координат атомов для аниона динитрамида $[\text{N}(\text{NO}_2)_2]^-$ приведен в листинге 10.

Листинг 10. Файл координат атомов для аниона динитрамида в xyz-формате

```

7
Dinitramide anion
N      -1.57397      2.01010      -0.06057
N      -0.43868      1.58643       0.49954
N       0.61886      2.23572       0.02190
O      -2.25481      1.09652      -0.52808
O      -1.85105      3.20784      -0.00858
O       1.00030      3.21935       0.65407
O       1.17194      1.67561      -0.92500

```

Что касается дополнительных параметров, то в большинстве случаев их как-либо изменять не требуется. К числу этих параметров относятся (1) выбор базиса для оптимизации геометрии молекулы и аниона, (2) выбор базиса для вычисления энергии молекулы и аниона, (3) выбор функционала для оптимизации геометрии молекулы и аниона, (4) выбор функционала для вычисления энергии молекулы и аниона, (5) заряд молекулы, (6) заряд аниона. Для изменения всех этих параметров необходимо внести изменения в скрипт данного WF (см. листинг 1).

При этом в блоке “step NormalGeom runs ORCA.DFT” задаются параметры, относящиеся к оптимизированной геометрии для нейтральной молекулы, в блоке “step NormalEnergy runs Gamess.DFT” - параметры для расчета полной энергии нейтральной молекулы, в блоках “step NormalAnion runs ORCA.DFT” и “step NormalAnion runs Gamess.DFT” – аналогичные параметры для расчета аниона.

В том случае, если исходное состояние диссоциирующей системы является не нейтральным, а заряженным, например, если оно имеет заряд +1, а диссоциированное состояние электронейтрально, то во всех четырех блоках нужно увеличить на 1 значение параметра charge.

7. ВЫХОДНЫЕ ДАННЫЕ

В принципе, пользователю доступны все четыре выходных файла, получающиеся в результате завершения каждого из четырех этапов решения задачи. Однако, для получения результата – величины рКа – необходимы только два выходных файла, выдаваемые пакетом GAMESS, содержащие полные энергии, рассчитанные для молекулы и для аниона.

На листинге 11 показан фрагмент выходного файла, выдаваемого пакетом GAMESS после расчета полной энергии нейтральной молекулы. Аналогичный фрагмент для аниона динитрамида приведен на листинге 12.

Листинг 11. Фрагмент выходного файла расчета полной энергии нейтральной молекулы динитрамида

----- RESULTS OF PCM CALCULATION -----		
FREE ENERGY IN SOLVENT = <PSI H(0)+V/2 PSI>	=	-465.4636837024 A.U.
INTERNAL ENERGY IN SOLVENT = <PSI H(0) PSI>	=	-465.4529274635 A.U.
DELTA INTERNAL ENERGY = <D-PSI H(0) D-PSI>	=	0.0000000000 A.U.
ELECTROSTATIC INTERACTION	=	-0.0107562388 A.U.
PIEROTTI CAVITATION ENERGY	=	0.0000000000 A.U.
DISPERSION FREE ENERGY	=	0.0000000000 A.U.
REPULSION FREE ENERGY	=	0.0000000000 A.U.
TOTAL INTERACTION (DELTA + ES + CAV + DISP + REP)	=	-0.0107562388 A.U.
TOTAL FREE ENERGY IN SOLVENT	=	-465.4636837024 A.U.

Листинг 12. Фрагмент выходного файла расчета полной энергии аниона динитрамида

----- RESULTS OF PCM CALCULATION -----		
FREE ENERGY IN SOLVENT = <PSI H(0)+V/2 PSI>	=	-465.0388814437 A.U.

INTERNAL ENERGY IN SOLVENT = <PSI H(0) PSI>	=	-464.9534981698 A.U.
DELTA INTERNAL ENERGY = <D-PSI H(0) D-PSI>	=	0.0000000000 A.U.
ELECTROSTATIC INTERACTION	=	-0.0853832739 A.U.
PIEROTTI CAVITATION ENERGY	=	0.0000000000 A.U.
DISPERSION FREE ENERGY	=	0.0000000000 A.U.
REPULSION FREE ENERGY	=	0.0000000000 A.U.
TOTAL INTERACTION (DELTA + ES + CAV + DISP + REP)	=	-0.0853832739 A.U.
TOTAL FREE ENERGY IN SOLVENT	=	-465.0388814437 A.U.

Используя формулы, приведенные в разделе 3.1 и проводя все вычисления в атомных единицах, отсюда легко получить: $D=0,424802$ а.е.;

$$(pK_a)_{расч.} = (D - 0.41345 - 2.5 * 0.00092797) / (2.3 * 0.00092797) = 4,231927 \text{ а.е.}$$

$$pK_a = 0.55(pK_a)_{расч.} - 6.9 = -4,57244.$$

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

МИТП	Многофункциональная инструментально-технологическая платформа
ПС	Прикладной сервис
DFT	Density Functional Theory, Теория функционала плотности (название группы методов)
PCM	Polarizable Continuum Model, Модель поляризуемого континуума

ПЕРЕЧЕНЬ ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. <http://www.thch.uni-bonn.de/tc/orca/>
2. http://www.thch.uni-bonn.de/tc/orca/index.php?option=com_docman&task=doc_download&gid=49&Itemid=26
3. <http://www.msg.chem.iastate.edu/gamess/documentation.html>
4. http://www.msg.chem.iastate.edu/gamess/GAMESS_Manual/input.pdf
5. http://avogadro.openmolecules.net/wiki/Main_Page

