

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Химический институт им. А.М. Бутлерова



подписано электронно-цифровой подписью

Программа дисциплины
Химическая информатика Б3.В.8

Направление подготовки: 020100.62 - Химия

Профиль подготовки: Органическая химия

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Маджидов Т.И.

Рецензент(ы):

Антигин И.С.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Антигин И. С.

Протокол заседания кафедры № ____ от "____" ____ 201____г

Учебно-методическая комиссия Химического института им. А.М. Бутлерова:

Протокол заседания УМК № ____ от "____" ____ 201____г

Регистрационный № 747814

Казань

2014

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) научный сотрудник, к.н. Маджидов Т.И. НИЛ Хемоинформатика и молекулярное моделирование Химический институт им. А.М. Бутлерова , Timur.Madzhidov@kpfu.ru

1. Цели освоения дисциплины

Целью освоения дисциплины "Химическая информатика" является подготовка к научно-исследовательской и педагогической деятельности, связанной с решением задач, стоящих перед наукой, посредством приложения современных средств хемоинформатики и методов молекулярного моделирования к химии во всех ее проявлениях. В результате освоения данной дисциплины должны быть сформированы представления о существующих методах молекулярного моделирования и хемоинформатики, их подходах, теоретическом обосновании, применимости к изучению различных явлений, о возможности использования современных ИТ-технологий и подходов для конкретных научных целей: предсказания биологической, фармацевтической и токсической активности соединений, исследования тонких особенностей их структуры, изучения и моделирования свойств, реакционной способности.

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б3.В.8 Профессиональный" основной образовательной программы 020100.62 Химия и относится к вариативной части. Осваивается на 4 курсе, 8 семестр.

Дисциплина "Химическая информатика" относится к вариативной части учебного профессионального цикла Б3. Она базируется на знаниях и умениях, выработанных при прохождении общих профессиональных курсов базовой части цикла "Органическая химия" (классификация и свойства органических соединений, механизмы органических реакций), "Физическая химия" (свойства молекул, механизмы реакций), "Строение вещества" (теория строения молекул и факторы, определяющие стабильность и реакционную способность соединений). Полученные при освоении дисциплины знания и умения облегчают освоение других курсов профессионального цикла "Современные инновационные методы в химии", "Электронная и пространственная структура органических соединений".

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-6 (профессиональные компетенции)	владеет навыками работы на современной учебно-научной аппаратуре при проведении химических экспериментов
ПК-7 (профессиональные компетенции)	имеет опыт работы на серийной аппаратуре, применяемой в аналитических и физико-химических исследованиях
ПК-8 (профессиональные компетенции)	владеет методами регистрации и обработки результатов химически экспериментов
ОК-6 (общекультурные компетенции)	использование основных законов естественнонаучных дисциплин в профессиональной деятельности, применение методов математического анализа и моделирования, теоретического и экспериментального исследования

В результате освоения дисциплины студент:

1. должен знать:

место химической информатики, квантовой химии и подхода силовых полей в теоретической химии, сферы применения и ограничения данных разделов теоретической химии; основные понятия, определения, методы и подходы, используемые в химической информатике; основы квантовой механики и квантовой химии;

2. должен уметь:

создавать, пополнять и использовать новые и существующие базы данных химических соединений и реакций; проводить квантохимическое и молекулярномеханическое моделирование различных явлений; подбирать необходимый метод химической информатики, квантовой химии или подхода силовых полей для решения конкретной задачи;

3. должен владеть:

методами квантовой химии, молекулярной механики и динамики, подхода QSAR, создания, пополнения и обработки баз данных; навыками по анализу химических баз данных для решения конкретных задач; навыками работы с современным программным обеспечением, предназначенным для проведения молекулярного моделирования; навыками работы с программным обеспечением и информационными ресурсами для работы с базами данных.

4. должен демонстрировать способность и готовность:

решать практические задачи и отвечать на вопросы, поставленные экспериментальными методами исследования, с использованием методов теоретической химии (хемоинформатики, методов силовых полей и квантовой химии); перейти от создания новых веществ и материалов методом "проб и ошибок" к созданию их с использованием прецизионного использования средств хемоинформатики и молекулярного моделирования

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины зачет в 8 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Направления теоретической химии: -квантовая химия - подход силовых полей - химическая информатика.	8	1	2	0	0	
2.	Тема 2. Методы силовых полей (молекулярная механика)	8	1	2	0	0	
3.	Тема 3. Поверхность потенциальной энергии и минимизация энергии	8	2	2	2	0	домашнее задание
4.	Тема 4. Конформационный поиск	8	3	2	2	0	
5.	Тема 5. Молекулярная динамика	8	4	2	2	0	творческое задание
6.	Тема 6. Основные принципы квантовой механики	8	5	2	0	0	
7.	Тема 7. Метод Хартри-Фока	8	5-6	2	4	0	домашнее задание
8.	Тема 8. Полуэмпирические методы и их применимость	8	7	2	2	0	
9.	Тема 9. Методы, учитывающие электронную корреляцию	8	8	2	2	0	
10.	Тема 10. Применение квантовой химии: расчеты структуры, свойств молекул и путей реакций.	8	9-10	2	4	0	творческое задание
11.	Тема 11. Хемоинформатика. Компьютерное представление молекул	8	10	2	0	0	
12.	Тема 12. Химические базы данных	8	11	0	2	0	
13.	Тема 13. Организация поиска в базах данных	8	11	2	0	0	

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
14.	Тема 14. Создание и управление базами данных	8	12	0	2	0	
15.	Тема 15. Молекулярные дескрипторы	8	12	2	0	0	
16.	Тема 16. Молекула как объект "химического пространства" Методы моделирования связи структура-свойство	8	13-14	2	4	0	тестирование
17.	Тема 17. Интеллектуальный анализ данных в химии	8	14	0	2	0	
18.	Тема 18. Виртуальный скрининг, основанный на структуре	8	15	2	0	0	
19.	Тема 19. Фармакофорный поиск	8	15-16	2	2	0	
20.	Тема 20. Создание лекарств с использованием средств хемоинформатики	8	16	0	2	0	контрольная работа
.	Тема . Итоговая форма контроля	8		0	0	0	зачет
	Итого			32	32	0	

4.2 Содержание дисциплины

Тема 1. Направления теоретической химии: -квантовая химия - подход силовых полей - химическая информатика.

лекционное занятие (2 часа(ов)):

Теоретическая химия как научная дисциплина. Методы молекулярного моделирования. Различие и комплементарность хемоинформатики, квантовой химии и методов силовых полей. Базовая молекулярная модель. Построение логического вывода. Представление химических объектов. Использование различных методов молекулярного моделирования для различных целей.

Тема 2. Методы силовых полей (молекулярная механика)

лекционное занятие (2 часа(ов)):

Силовые поля. Основные применения и ограничения. Параметры силовых полей. Учет силовых констант валентных взаимодействий. Учет невалентных взаимодействий. Наиболее распространенные силовые поля и цели их использования. Периодические граничные условия.

Тема 3. Поверхность потенциальной энергии и минимизация энергии

лекционное занятие (2 часа(ов)):

Гиперповерхность потенциальной энергии. Локальный и глобальный минимум. Характеристики различных точек на поверхности потенциальной энергии. Первая производная, градиент и гессиан энергии. Понятие минимизации энергии. Методы минимизации нулевого, первого и второго порядка. Линейный поиск. Метод скорейшего спуска. Метод сопряженных градиентов. Метод Ньютона-Рафсона. Квазиньютоновские методы.

практическое занятие (2 часа(ов)):

Обучение способам минимизации энергии с использованием средств пакета VegaZZ. Сравнение различных методов минимизации.

Тема 4. Конформационный поиск

лекционное занятие (2 часа(ов)):

Конформационный поиск. Сравнение различных методов конформационного поиска. Систематический конформационный поиск. Стохастический конформационный поиск. Метод Монте-Карло. Генетический алгоритм. Алгоритм пчелиного роя. Генерация конформеров с использованием правил и заранее определенных данных. Преимущества и недостатки. Наиболее распространенные генераторы конформеров.

практическое занятие (2 часа(ов)):

Генерация конформеров с использованием средств программ VegaZZ, ChemAxon. Сравнение наборов конформеров.

Тема 5. Молекулярная динамика

лекционное занятие (2 часа(ов)):

Молекулярная динамика. Ньютоновская, лагранжевская и гамильтоновская механика. Типы ансамблей. Интеграторы и алгоритмы интегрирования траекторий. Контроль температуры и давления при моделировании.

практическое занятие (2 часа(ов)):

Генерация и анализ молекулярно-динамических траекторий движения комплексов молекул.

Тема 6. Основные принципы квантовой механики

лекционное занятие (2 часа(ов)):

Корпускулярно-волновой дуализм. Предпосылки возникновения квантовой механики. Волновая функция. Понятие оператора. Требования к операторам. Важнейшие операторы. Конструирование оператора. Понятие физически наблюдаемой характеристики. Основные постулаты квантовой механики. Собственное значение и собственная функция оператора. Среднее значение оператора. Уравнение Шредингера. Спин электрона. Принцип суперпозиции. Антисимметрия волновой функции.

Тема 7. Метод Хартри-Фока

лекционное занятие (2 часа(ов)):

Представление волновой функции в методе Хартри. Детерминант Слэйтера. Молекулярные орбитали. Использование детерминанта Слэйтера в решении уравнения Шредингера. Приближение МО ЛКАО. Атомные орбитали. базисные функции. Одно- и двухэлектронные интегралы. Кулоновский и обменный интеграл. Матрица Фока. Матрица перекрывания. Поиск значений коэффициентов перед атомными орбиталями в молекулярной орбитали. Итеративный способ самосогласования поля.

практическое занятие (4 часа(ов)):

Оптимизация геометрии с использованием метода Хартри-Фока.

Тема 8. Полуэмпирические методы и их применимость

лекционное занятие (2 часа(ов)):

Приближения, используемые в создании полуэмпирических методов. Приближение нулевого дифференциального перекрывания. Назначение. Приближение пренебрежения двуцентровым дифференциальным перекрыванием, частичного пренебрежения дифференциальным перекрыванием. Модифицированные методы ПДДП (NDDO). Методы MNDO, AM1, PM3.

практическое занятие (2 часа(ов)):

Проведение расчетов структуры и характеристик молекул полуэмпирическими методами.

Тема 9. Методы, учитывающие электронную корреляцию

лекционное занятие (2 часа(ов)):

Понятие электронной корреляции. Динамическая и статическая электронная корреляция. Теория функционала плотности. Понятие функционала. Метод Кона-Шама. Виды функционалов: локальные, градиент-корректированные, гибридные. Пост-хартрифоковские методы. Методы теории возмущений. Методы конфигурационного взаимодействия. Метод многоконфигурационного взаимодействия.

практическое занятие (2 часа(ов)):

Проведение расчетов геометрии и энергии комплексообразования методами теории функционала плотности.

Тема 10. Применение квантовой химии: расчеты структуры, свойств молекул и путей реакций.

лекционное занятие (2 часа(ов)):

Использование методов квантовой химии для расчета: энергетических характеристик (относительная энергия конформеров, таутомеров, энергия, энталпия и свободная энергия реакции), спектральных характеристик (колебательные спектры, электронные спектры, спектры ЯМР), заряды, электростатический потенциал. Расчет переходных состояний и интермедиатов. Расчет путей реакций.

практическое занятие (4 часа(ов)):

Расчеты свойств молекул. Сравнение качества различных методов расчета.

Тема 11. Хемоинформатика. Компьютерное представление молекул

лекционное занятие (2 часа(ов)):

Хемоинформатика как научная дисциплина. Хемоинформатика и связанные с ней дисциплины. Хемоинформатика и хемометрика. Хемоинформатика и биоинформатика. Представление молекул. Линейные представления (WLN, SMILES, InChI). Графовые представления: векторные, матричные и табличные представления. Структуры Маркуша. Трехмерные представления: координатные представления, молекулярные поверхности, молекулярные формы. Форматы файлов MDL (MOL, SDF, RXN, RDF), CML, PDB.

Тема 12. Химические базы данных

практическое занятие (2 часа(ов)):

Редактирование молекул и сохранение их в различных форматах. Конвертация между представлениями.

Тема 13. Организация поиска в базах данных

лекционное занятие (2 часа(ов)):

Общие сведения о базах данных. Классификация баз данных. Структурный поиск в химических базах данных. Виды поиска, основанного на структуре: структурный, подструктурный и суперструктурный поиск. Поиск по схожести. Алгоритмы осуществления поисков.

Тема 14. Создание и управление базами данных

практическое занятие (2 часа(ов)):

Создание баз данных с использованием инструментов ChemAxon.

Тема 15. Молекулярные дескрипторы

лекционное занятие (2 часа(ов)):

Молекулярные дескрипторы. Классификация дескрипторов. Топологические индексы. Двумерные автокорреляции. 3D дескрипторы. Фрагментные дескрипторы. Фармакофорные дескрипторы. Константы заместителей. Физико-химические дескрипторы. Квантохимические дескрипторы. Дескрипторы молекулярных полей. Дескрипторы молекулярного подобия.

Тема 16. Молекула как объект "химического пространства" Методы моделирования связи структура-свойство

лекционное занятие (2 часа(ов)):

Объекты химического пространства. Отношения сходства между объектами химического пространства. Уровни отношения сходства. Отношения сходства для различных математических представлений химических объектов. Расстояния и метрики. Дескрипторы как характеристики положения молекул в химическом пространстве. Моделирование структура-свойство (QSPR/QSAR/SAR). Классификационные и регрессионные методы для создания моделей, связывающих структуру и свойства молекул. Валидация моделей. Чистка данных.

практическое занятие (4 часа(ов)):

Создание классификационных и регрессионных моделей для связи структуры молекулы с ее свойствами с использованием программы Weka и ресурса OChem

Тема 17. Интеллектуальный анализ данных в химии

практическое занятие (2 часа(ов)):

Элементы теории машинного обучения. Основные принципы машинного обучения. Основные принципы машинного обучения. Принципы построения моделей в классической статистике. Функция правдоподобия и метод наименьших квадратов (МНК). Критерии оценки качества регрессионных моделей. Критерии оценки качества классификационных моделей. Методы машинного обучения. Множественная линейная регрессия. Построение линейных моделей с регуляризацией. Метод главных компонент (PCA). Метод частичных наименьших квадратов (PLS). Искусственные нейронные сети. Деревья принятия решений. Метод опорных векторов.

Тема 18. Виртуальный скрининг, основанный на структуре

лекционное занятие (2 часа(ов)):

Виртуальный скрининг по структуре биомолекулы. Подготовка структуры биомолекулы. Источники трехмерной структуры белков. Построение пространственной структуры молекулы белка методами сравнительного моделирования. Построение пространственной структуры молекулы белка методами ab initio. Скрининг по структуре. Докинг. Жесткий докинг. Гибкий по лиганду докинг. Скоринг. Виды скоринг-функций.

Тема 19. Фармакофорный поиск

лекционное занятие (2 часа(ов)):

Фармакофоры. Виды фармакофорных центров. Топологические фармакофоры. Трехмерные фармакофоры. Фармакофорный поиск. Методы выравнивания фармакофоров. Скрининг по фармакофорам.

практическое занятие (2 часа(ов)):

Виртуальный скрининг с использованием методов фармакофорного поиска.

Тема 20. Создание лекарств с использованием средств хемоинформатики

практическое занятие (2 часа(ов)):

Дизайн лекарств. Принципы комплементарности в создании лекарственных препаратов. Пептидомиметики. Полифармакология веществ. Использование сетей пептид-пептидного взаимодействия в дизайне лекарственных препаратов. Предсказание ADME/Tox свойств. Фильтры схожести с лекарством. Правило Липинского. Типичный алгоритм виртуального скрининга для создания лекарственных препаратов.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
3.	Тема 3. Поверхность потенциальной энергии и минимизация энергии	8	2	подготовка домашнего задания	6	домашнее задание

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
5.	Тема 5. Молекулярная динамика	8	4	подготовка к творческому экзамену	8	творческое задание
7.	Тема 7. Метод Хартри-Фока	8	5-6	подготовка домашнего задания	6	домашнее задание
10.	Тема 10. Применение квантовой химии: расчеты структуры, свойств молекул и путей реакций.	8	9-10	подготовка к творческому экзамену	8	творческое задание
16.	Тема 16. Молекула как объект "химического пространства" Методы моделирования связи структурно-свойство	8	13-14	подготовка к тестированию	8	тестирование
20.	Тема 20. Создание лекарств с использованием средств хемоинформатики	8	16	подготовка к контрольной работе	8	контрольная работа
Итого					44	

5. Образовательные технологии, включая интерактивные формы обучения

- компьютерные презентации лекций;
- проведение дискуссий;
- работа с современным программным обеспечением;
- презентация фильмов по работе тех или иных технологий;
- внеаудиторная самостоятельная работа с современным программным обеспечением с написанием отчетов по выполненной работе;
- встречи с представителями российских и зарубежных компаний, работающих в области создания веществ с заданными свойствами;
- мастер-классы российских и зарубежных специалистов в области хемоинформатики.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Направления теоретической химии: -квантовая химия - подход силовых полей - химическая информатика.

Тема 2. Методы силовых полей (молекулярная механика)

Тема 3. Поверхность потенциальной энергии и минимизация энергии

домашнее задание , примерные вопросы:

Компьютерное моделирование неорганических, органических и элементоорганических соединений с использованием силовых полей. Использование техники минимизации энергии системы для предсказания структуры. Сравнение с экспериментом.

Тема 4. Конформационный поиск

Тема 5. Молекулярная динамика

творческое задание , примерные вопросы:

Моделирование динамики и структуры биологических макромолекул и органических полимеров с использованием молекулярной динамики.

Тема 6. Основные принципы квантовой механики

Тема 7. Метод Хартри-Фока

домашнее задание , примерные вопросы:

Компьютерное моделирование структуры малых молекул с использованием метода Хартри-Фока. Минимизация энергии, сравнение расчетов с экспериментом. Анализ электронного распределения - характеристик молекулярных орбиталей, зарядов.

Тема 8. Полуэмпирические методы и их применимость

Тема 9. Методы, учитывающие электронную корреляцию

Тема 10. Применение квантовой химии: расчеты структуры, свойств молекул и путей реакций.

творческое задание , примерные вопросы:

Моделирования характеристик молекул с использованием средств квантовой химии. Выбор оптимального метода расчета. Объяснение экспериментально наблюдаемых фактов.

Тема 11. Хемоинформатика. Компьютерное представление молекул

Тема 12. Химические базы данных

Тема 13. Организация поиска в базах данных

Тема 14. Создание и управление базами данных

Тема 15. Молекулярные дескрипторы

Тема 16. Молекула как объект "химического пространства" Методы моделирования связи структура-свойство

тестирование , примерные вопросы:

Тестирование по темам 11-16, включающее вопросы, связанные с различными видами представлений химических объектов, структурных поисков, дескрипторного представления молекул, характеристик объектов в химическом пространстве

Тема 17. Интеллектуальный анализ данных в химии

Тема 18. Виртуальный скрининг, основанный на структуре

Тема 19. Фармакофорный поиск

Тема 20. Создание лекарств с использованием средств хемоинформатики

контрольная работа , примерные вопросы:

Подготовка к контрольной работе по темам 11-20

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

Примерный вариант задания к теме 3:

Необходимо провести моделирование молекулы диаланина:

1. Нарисуйте различные конформации молекулы диаланина (различные виды а-спиралей, β -складчатого листа, спиралей коллагена) в соответствии с приведенными значениями углов ϕ , ψ и ω .

2. Рассчитайте энергию каждого конформера молекулы в данной точке (без оптимизации Calculate>AMMP>Minimization, выбрать Single point). Энергию записать в таблицу ниже - в графу Etot

(исх). Проведите минимизацию энергии с использованием алгоритма Quasi-Newton в программе VegaZZ,

энергию запишите в графу Etot (конеч).

3. Проведите анализ результатов. Ответьте на следующие вопросы:

- какой из исходных конформер наиболее устойчив? В каком порядке изменяется стабильность конформеров?
- какой из конечных конформеров наиболее устойчив? В каком порядке изменяется стабильность полученных конформеров?
- сильно ли изменяется геометрия при оптимизации?
- можно ли сделать вывод относительно того, как зависит энергия от величин углов и гош-транс ориентации групп в конформерах?

Примерный вариант задания к теме 5:

Проведите молекулярно-динамическое моделирование молекулы диаланина, погруженного в растворитель (вода).

1. Нарисуйте молекулу диэтилового эфира в программе VegaZZ. Погрузите его в окружение молекул воды с использованием средств программы. Размер ячейки - 10 ангстрем.

2. Проведите молекулярнодинамическое моделирование системы с использованием программы NAMD и силового поля CHARMM. Подберите параметры моделирования и обоснуйте ваш выбор.

3. Проведите анализ результатов. Ответьте на следующие вопросы:

- какая конформация диаланина является наиболее устойчивой в воде?
- в каких диапазонах изменяются углы ϕ , ψ и ω в ходе моделирования? Относительно каких связей вращение заторможено и относительно каких - происходит легко?
- какие атомы наиболее склонны к образованию водородных связей?
- какое число водородных связей в среднем образует молекула диаланина?

Примерный вариант задания к теме 7:

Проведите оптимизацию геометрии молекулы воды в базисах cc-pvdz, cc-pvtz, cc-pvqz.

Сравните полученные энергии друг с другом. К одинаковым ли изменениям энергии приводит увеличение базисного набора? Сравните геометрию полученных структур с данными микроволновой спектроскопии. Можно ли сказать, что увеличение базисного набора приводит к улучшению описания геометрии? Сравните симметрию, относительные энергии молекулярных орбиталей и вклады в них от атомных орбиталей с теми, что получаются при качественном предсказании молекулярных орбиталей (приведена на рисунке).

Примерное задание к теме 10:

Согласно экспериментальным данным, химические сдвиги протонов этанола в различных малополярных растворителях расположены области:

метильная группа - 1.25 м.д. (CDCl_3), 0.96 м.д. (C_6D_6)

метиленовая группа - 3.72 м.д. (CDCl_3), 3.34 м.д. (C_6D_6)

химические сдвиги углеродов - в области:

метильная группа - 18.41 м.д. (CDCl_3), 18.72 м.д. (C_6D_6)

метиленовая группа - 58.28 м.д. (CDCl_3), 57.86 м.д. (C_6D_6)

Переберите несколько методов расчетов химических сдвигов - несколько методов DFT и базисных наборов. Однаково ли хорошо воспроизводятся химические сдвиги углерода и водорода? Как влияет базис на точность расчета химических сдвигов? Проведите расчет с использованием базисного набора, состоящего только из примитивных гауссианов. Насколько точно воспроизводятся характеристики? Предложите способ как можно более точного расчета химических сдвигов за кратчайшее время.

Примерный вариант теста по темам 11-16:

1. Какие из указанных SMILES соответствует молекуле аспирина?

a. CC(=O)Oc1ccccc1C(O)=O

- b. c1c(C(=O)=O)cc(OC(OC)C)ccc1
- c. OC(=O)c(cccc1)c1OC(=O)C
- d. c1(C(=O)O)cccc1OC(=O)C

2. Какие InChI для приведенной молекулы гуанина соответствуют молекуле и являются стандартными?

- a. InChI=1/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)/f/h8,10H,6H2
- b. InChI=1S/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)
- c. InChI=1S/C6H6N5O2/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)
- d. InChI=1/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)/f/h7,9H,6H2

3. Какой SMARTS запроса будет определять выделенную подструктуру в приведённой молекуле? Атомы водорода не принимать во внимание.

- a. N~*~*~N
- b. NcccN
- c. [#7]ccc[#7]
- d. [NH2]aaa[NH2]

4. Какая из приведенных SMILES удовлетворяет приведенной структуре Маркуша?

- a. OCCc1c(C)cccc1
- b. OCCCCCCCc1cc(C(=O)O)ccc1
- c. OCCCc1ccc(C(C)=O)cc1
- d. OCCCc1ccc(COC=O)cc1

5. Какое из приведенных отнесений отмеченных фармакофорных центров 1,2 и 3 является наиболее полным и корректным (один ответ)? Обозначения: N - negative charge, P - positive charge, H - hydrophobe, Ar - aromatic ring, A - H-acceptor, D - H-donor.

- a. 1: N; 2: D; 3: P;
- b. 1: N; 2: A, D; 3: P, D;
- c. 1: N, A; 2: A, D; 3: P, D, Ar;
- d. 1: N, A; 2: A, D, N; 3: P, A, D, Ar;

6. Каким из приведенных молекул (1,2,3, 4) удовлетворяет приведенный трехточечный топологический фармакофор P? Обозначения: P - positive charge, A - H-acceptor, D - H-donor. Расстояния являются топологическими.

- a. 1
- b. 2
- c. 3
- d. 4

7. Каким из приведенных молекул (1,2,3, 4) соответствует данная структура Маркуша?

- a. 1
- b. 2
- c. 3
- d. 4

8. Какая из указанных структур (1,2,3,4), содержащихся в базе, будет выдаваться в результате поиска по субструктуре M? Атомы водорода не принимаются во внимание.

- a. 1
- b. 2
- c. 3
- d. 4

9. Какая из указанных структур (1,2,3,4), содержащихся в базе, будет выдаваться в результате поиска по суперструктуре M?

- a. 1
- b. 2
- c. 3
- d. 4

10. Две структуры задаются указанными ниже битовыми строками. Какой будет индекс схожести Танимото между данными структурами?

Mol 1 1 0 0 0 1 1 1 0 1 1

Mol 2 0 1 0 0 1 0 1 1 0 1

- a. 6/11
- b. 5/8
- c. 3/10
- d. 3/8

11. Какое из приведенных выражений содержит формулу для вычисления индекса схожести Тверского? а - число активных бит в одной молекуле, b - число включенных бит в другой молекуле, c - число бит, которые являются активными в обеих молекулах.

- a. $c/(a+b-c)$
- b. $2c/(a+b)$
- c. $(a+b-2c)/(a+b-c)$
- d. $c/(c+\alpha(a-c)+\beta(b-c))$

12. В какой из приведенных баз данных можно найти информацию, характеризующую прочность связывания данного химического соединения с различными белками?

- a. CAS
- b. PubChem
- c. ChEMBL
- d. ZINC

13. Какую информацию о соединении можно найти в базе ChemSpider?

- a. Химическая структура
- b. Информация об испытании данного соединения на bioassay
- c. Кристаллическая структура молекулы
- d. Индекс LASSO, характеризующий насколько данная молекула подходит для связывания с активными центрами различных ферментов

14. Какие этапы входят в процесс осуществления поиска по структуре?

- a. Стандартизация соединения
- b. Генерация хэш-кода
- c. Поиск молекулы с помощью скринов
- d. Поиск индекса схожести данного соединения с другими соединениями базы

Примерный вариант контрольной работы к теме 20:

1. На основании матрицы связей нарисовать молекулу. Дать ей имя с использованием SMILES. Создать на ее основе матрицу расстояний (2 балла)

1 2 3 4 5 6 7 8 9 10 11 12

N 1 0 0 0 2 0 0 0 0 0 0

C 2 0 0 0 0 0 0 0 0 0

C 1 0 0 0 0 0 0 0 0

C 2 0 0 0 0 0 0 0 0

C 1 1 0 0 0 0 0 0

C 0 0 0 0 0 0 0 0

C 1 0 0 1 0

C 1 0 0 0

C 1 0 0

C 1 0

N 1

C

2. Для указанных пяти соединений с использованием приведенных фрагментов и расстояние Сергела (=1-Танимото) проведите кластеризацию с использованием метода single-link и complete-link. (3 балла)

3. С использованием cell-based метода для анализа разнообразия молекул отберите 7 наиболее разнообразных молекул из базы данных 1. Дополните базу данных 1 минимальным числом соединений из базы данных 2, чтобы создать максимально разнообразный набор.

Рекомендуемые диапазоны корзин LogP: 6; MW : 0-250, 250-500, > 500.

База данных 1

logP MW

mol 1 -1.32 664

mol 2 3.52 301

mol 3 1.39 154

mol 4 5.28 384

mol 5 -1.04 473

mol 6 6.17 256

mol 7 1.21 260

mol 8 4.9 450

mol 9 -0.1 304

mol 10 8.75 625

База данных 2

logP MW

mol 11 10.3 390

mol 12 -2.84 228

mol 13 -0.69 74

mol 14 0.05 508

mol 15 0.4 79

mol 16 3.19 232

mol 17 4.63 586

(2 балла)

4. Проведите нормализацию дескрипторов для соединений из базы 1, приведенных ниже, на нулевое среднее и на единичную вариацию. Найдите евклидовы расстояния между молекулами и проведите поиск 4x наиболее различающихся соединений с использованием алгоритмов MaxMin и MaxSum. В качестве исходного возьмите соединение наиболее отличающееся от других. (4 балла)

logP MW

mol 1 -1.32 664

mol 2 3.52 301

mol 4 5.28 384

mol 5 -1.04 473

mol 6 6.17 256

mol 9 -0.1 304

5. Полученную в предыдущем задании матрицу используйте для проведения неиерархической кластеризации по методу Ярвиса-Патрика с использованием следующих параметров:

- a) m=2, k=1
 - b) m=3, k=2
- (3 балла)

Билеты к зачету:

Билет ♦ 1

1. Линейное представления SMILES. Правила. Представление стереоизомеров, таутомеров, ароматических структур. Уникальные, абсолютные, изомерные и реакционные SMILES.
2. Предобработка данных. Химическая предобработка: отбор данных и стандартизация. Математическая предобработка: стандартизация, шкалирование, нормализация.

Билет ♦ 2

1. Представления химической реакций. Представление реакции как набора реагентов и продуктов. Представление реакции Фуджита. Представление Угги-Дугуджи. Представления реакций как разности продуктов и реагентов. Представления реакций как характеристик реакционного центра. Атом-атомное отображение.
2. Характеристики качества моделей. Характеристики регрессионных моделей: R, R², Q², REC. Характеристики классификационных моделей: матрица невязок, ROC, TPR, FPR, BA, AUC и др.

Билет ♦ 3

1. Поиск подструктуре и суперструктуре. Основные алгоритмы поиска. Использование скринов. Изоморфное вложение.
2. Способы борьбы с переобучением: Y-scrambling, скользящий контроль, контроль по блокам (n-fold cross-validation). Внешний и внутренний тестовый набор. Требование к внешнему тестовому набору. Трехвыборочный подход.

Билет ♦ 4

1. Дизайн комбинаторных библиотек: подходы к проблеме создания оптимальной комбинаторной библиотеки и методы решения.
2. Домен применимости модели. Понятие. Способы оценки домена применимости: контроль дескрипторов, z-kNN, 1-SVM.

Билет ♦ 5

1. Пост-хартрифоковские методы: метод конфигурационного взаимодействия, многоконфигурационного взаимодействия.
2. Использование биоизостеризма химических групп для создания лекарственных препаратов.

Билет ♦ 6

1. Кластеризация. Виды кластеризации. Иерархическая и неиерархическая. Агglomerативная и дивизимная.
2. Слабый и сильный метод обучения (weak, strong learner). Улучшение качества моделей. Консенсусные подходы. Ансамблевое обучение. Методы: бэггинг, бустинг, стэкинг, использование случайного поднабора. Бутстрэп.

Билет ♦ 7

1. Линейное представления InChI. Правила. Представление стереоизомеров, таутомеров, ароматических структур. Стандартные InChI.

2. Метод опорных векторов (SVM). Геометрический и функциональный зазор. Постановка задачи машинного обучения. Решение задачи с помощью обобщенного метода множителей Лагранжа. Решение задачи классификации в случае линейно неразделимых выборок (C-SVM).

Билет ♦ 8

1. Химическое пространство. Виды химического пространства: пространство на дескрипторах, на графах, на функциях, облако точек.
2. Линейное и нелинейное разделение в SVM. Понятие о ядре Мерсера (kernel). Требования к ядру. Преобразование пространства с помощью ядра. Спрямляющее пространство. Виды ядер. Подбор ядра.

Билет ♦ 9

1. Трехмерный фармакофорный поиск. Двухступенчатый алгоритм поиска. Фармакофорное отображение. Основные подходы к отображению (с полным перебором конформеров, с использованием клики, генетического алгоритма).
2. Навигация в химическом пространстве дескрипторов. Проекция химического пространства для визуализации. Алгоритмы проекции (общая идея): PCA, SOM, GTM.

Билет ♦ 10

1. Навигация в химическом пространстве графов. Виды. Скаффолды и каркасы. Скаффолды Бемиса-Мурко. Подход дерева скаффолдов Ертля с соавт.
2. Метод множественной линейной регрессии (MLR). Отбор дескрипторов: необходимость отбора, способы (жадные, стохастические методы). Проблема смещенностя выбора дескрипторов и способ борьбы с нею. Снижение склонности к переобучению.

7.1. Основная литература:

1. Чмутова, Г.А. Аспекты связи "Строение - реакционная способность": учебное пособие / Г. А. Чмутова; Казан. (Приволж.) федер. ун-т, Хим. ин-т.?Казань: [Казанский (Приволжский) федеральный университет], 2010.?93 с.
2. Введение в хемоинформатику. Компьютерное представление химических структур: учебное пособие / Т. И. Маджидов [и др.].?Казань: Казанский университет, 2013.?173 с.
3. Гвоздева В. А. Базовые и прикладные информационные технологии: Учебник / В.А. Гвоздева. - М.: ИД ФОРУМ: НИЦ ИНФРА-М, 2014. - 384 с.
<http://znanium.com/bookread.php?book=428860>
4. Федотова Е. Л. Прикладные информационные технологии: Учебное пособие / Е.Л. Федотова, Е.М. Портнов. - М.: ИД ФОРУМ: НИЦ ИНФРА-М, 2013. - 336 с. Режим доступа:
<http://znanium.com/bookread.php?book=392462>

7.2. Дополнительная литература:

1. Маджидов Т.И. Хемоинформатика и молекулярное моделирование: дистанционный курс для студентов бакалавриата и магистратуры направления подготовки: 020100 "Химия" [Электронный ресурс]. Площадка "Зилант" СУО КФУ, 2013. // <http://zilant.kpfu.ru/course/view.php?id=376>
2. Каплан И. Г. Межмолекулярные взаимодействия. Физическая интерпретация, компьютерные расчеты и модельные потенциалы. Пер. с англ. Москва Бином. Лаборатория знаний, 2012. 394 с. http://e.lanbook.com/books/element.php?pl1_id=8690
3. Седых А.Е., Галкина И. В., Галкин В. И. Программа "ХСНЕМ" - использование фрагментов химической структуры для поиска и моделирования химических и биологических свойств.// Ученые записки КФУ. Естественные науки. 2009. N1.
http://libweb.ksu.ru/e-journals/1815-6169/2009/151_1/151_1_est_8.pdf

4. Татаринов Д.А., Немтарев А.В. Онлайн поисковые системы научной информации. / учебно-методическое пособие. - Казань: Казанский (Приволжский) федеральный университет. 30 с. Подробности: http://kpfu.ru/publication?p_id=72662

7.3. Интернет-ресурсы:

ChemAxon, tools for chemoinformatics - <http://www.chemaxon.com/>

Virtual Computational Chemistry Laboratory - <http://www.vcclab.org/>

Программа Firefly - <http://classic.chem.msu.su/gran/firefly/index.html>

Программа ISIDA-QSPR - <http://www.vpsolovev.ru/programs/>

Электронный образовательный ресурс ?Хемоинформатика и молекулярное моделирование? площадки "Зилант" системы дистанционного обучения Казанского (Приволжского) федерального университета - <http://zilant.kfu-elearning.ru/course/view.php?id=376>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Химическая информатика" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

- Компьютерный проектор;
- Компьютерный класс на 10 машин с предустановленным ПО.

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 020100.62 "Химия" и профилю подготовки Органическая химия .

Автор(ы):

Маджидов Т.И. _____
"___" 201 ___ г.

Рецензент(ы):

Антипов И.С. _____
"___" 201 ___ г.