

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт «Умные материалы и технологии»

УТВЕРЖДЕНО:
Директор Института «Умные
материалы и технологии»
И.А. Курзина

Рабочая программа дисциплины

Хемоинформатика

по направлению подготовки

27.03.05 Инноватика

Направленность (профиль) подготовки:

Tomsk International Science Program, с профессиональным модулем Молекулярная инженерия / Molecular Engineering

Форма обучения

Очная

Квалификация

Инженер

Год приема

2024

СОГЛАСОВАНО:
Руководитель ОП
И.А. Курзина

Председатель УМК
Г.А. Воронова

Томск – 2024

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-3. Способен применять современные информационные компьютерные технологии, обрабатывать и использовать новую информацию в предметной области.

ПК-2. Способен решать профессиональные задачи на основе знаний в сфере биотехнологии и молекулярной инженерии на основе знаний естественных, математических и технических наук, а также математических методов и моделей.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

РООПК-3.1. Знает принципы работы современных информационных компьютерных технологий, программ и сред.

РООПК-3.2. Умеет осваивать и применять современные информационные компьютерные технологии, программы и среды для обработки и получения информации в профессиональной сфере.

РОПК-2.1. Знает существующие подходы к решению профессиональных задач, в том числе на основе математических методов и моделей.

2. Задачи освоения дисциплины

– сформировать представления о предмете хемоинформатики, ее основных понятиях, методах и подходах, а также возможности использования ее методов и подходов для научно-практических целей;

– научить магистрантов использовать средства хемоинформатики для предсказания структуры соединения с требуемыми биологическими, химическими и физико-химическими свойствами;

– освоить методы хемоинформатики, требующиеся для решения тех или иных задач в химии;

– узнать способы представления химических данных, методы осуществления поиска в химических базах данных; основные химические базы данных, используемые в различных научных целях, и методы работы с ними.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Пятый семестр, экзамен.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Квантовая химия», «Физическая химия», «Органическая химия», «Строение вещества», полученные в рамках обучения по программе бакалавриата или специалитета.

6. Язык реализации

Английский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 6 з.е., 216 часов, из которых:

– лекции: 30 ч.

– лабораторные работы : 60 ч.;

– практические занятия : 0 ч.

– семинарские занятия: 20 ч.

в том числе практическая подготовка: 80 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Введение в дисциплину

Введение в дисциплину. Основные проблемы химии. Прямая и обратная задачи моделирования. Их решение. Предназначение хемоинформатики. Определение хемоинформатики. Хемоинформатика как научная дисциплина. Хемоинформатика как дисциплина теоретической химии. История хемоинформатики.

Тема 2. Представление химических объектов

Представление молекул. Типичные представления молекул в химии (структурная формула, химическая формула, тривиальное имя). Особенности представления в хемоинформатике, требования к представлениям. Виды представлений. Линейные представления (имена, WLN, SMILES, SLN, InChI). Представление молекулярных графов. Битовые строки (структурные ключи, отпечатки пальцев, хэшированные отпечатки пальцев). Матричное представление, виды матриц. Табличное представление. Трехмерные представления. Координаты атомов. Поверхности. Виды поверхностей (ван-дер-ваальсова поверхность, поверхность Коннолли, доступная растворителю поверхность, поверхность исключенного растворителя, поверхность полости фермента, поверхность изоплотности, раскрашенные поверхности). Молекулярная формы. Структуры Маркуша. Типичные форматы файлов (MDL, Sybyl, PDB). Конвертация между представлениями. Конверсия структура-имя и имя-структура. Конверсия структуры в линейные представления. 2D-3D конвертация. Представления реакций. Типичное представление реакций. Представление реакции как набора реагентов и продуктов. Представления реакций как характеристик реакционного центра. Представления реакций как разности продуктов и реагентов. Представление реакции Фуджита. Представление Угги-Дугуджи. Представления химических структур - ручное создание и интерпретация представлений SMILES, InChI. Редактирование структуры молекул с использованием редакторов. Создание файлов, содержащих требуемое представление молекул. Поиск в базе данных по SMARTS. Редактирование реакций. Поиск атом-атомного соответствия в реакциях. Создание файлов, содержащих различные представления реакций.

Тема 3. Химические базы данных

Химические базы данных. Типы баз. Базы молекул, спектров, белков, кристаллографические, биомолекулы. Виды поиска в химических базах данных. Поиск по структуре, подструктуре, суперструктуре и по молекулярному сходству в базах данных. Основные алгоритмы поиска. Использование скринов. Рекурсивный подход. Ульмановский подход. Поиск в 3D базах данных. Фармакофоры. Фармакофорный поиск. Создание и управление компьютерных баз данных химических соединений, реакций и смесей.

Тема 4. Молекулярное разнообразие

Дизайн библиотек данных. Использование для виртуального скрининга и для высокопроизводительного скрининга. Теоретическая комбинаторная химия. Разбросанные и сфокусированные библиотеки. Генерация структур. RECAP. Fragmenter. Кластеризация молекул. Иерархические подходы. Неиерархические подходы. Отбор молекул без кластеризации. Навигация в химическом пространстве молекулярных графов и основанные на скаффолдах методы кластеризации объектов в химическом пространстве.

Сравнение библиотек химических соединений, кластеризация объектов и отбор молекул в выборку на основании этого. Навигация в химическом пространстве с использованием метода GTM.

Тема 5. Молекулярные дескрипторы

Дескрипторы. Определение и использование дескрипторов. Роль дескрипторов в хемоинформатике. Многообразие дескрипторов. Классификация дескрипторов по функциональности. Физико-химические дескрипторы. Топологические индексы. Трехмерные. Фрагментные дескрипторы. Фармакофорные дескрипторы. Константы заместителей. Квантово-химические дескрипторы. Дескрипторы молекулярных полей. Дескрипторы молекулярного подобия. Расчет дескрипторов для молекул. Создание входных файлов для анализа связи структуры с реакционной способностью.

Тема 6. Моделирование «структура-свойство»

История моделирования «структура-свойство» SAR/QSAR/QSPR. Классический QSAR (методы Ганча, Фри-Вильсона). SAR/QSAR/QSPR на дескрипторах. Современные подходы. Методы машинного обучения. Интеллектуальный анализ данных в хемоинформатике. Задачи машинного обучения. Методы. MLR, гребневая регрессия, PLS, kNN, искусственные нейронные сети, SVM, решающие деревья, наивный Байес, гауссовы процессы. Использование методов машинного обучения. Валидация и кросс-валидация. ROC кривые. Предобработка данных. Химическая предобработка: отбор данных и стандартизация. Математическая предобработка: стандартизация, шкалирование, нормализация. Случайная корреляция и борьба с ней. Консенсусные подходы. Область применимости. 3D QSAR, основанный на пространственном выравнивании. Методы пространственного выравнивания. CoMFA, CoMSIA, Grid. Методы 3D QSAR, независимые от выравнивания. Grind. Общее понятие об nD QSAR. История моделирования «структура-свойство» SAR/QSAR/QSPR. Отбор дескрипторов, построение моделей структура-свойство и сравнение качества моделей.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости и оценивания отчетов по выполненным практическим работам, и фиксируется в форме контрольной точки не менее одного раза в семестр. При выполнении всех практических заданий студент допускается к сдаче зачета.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» – <https://www.tsu.ru/sveden/education/eduop/>

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет по курсу «Хемоинформатика» проводится в форме устного опроса студентов. Результаты зачета определяются оценками «зачтено» или «не зачтено».

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» – <https://www.tsu.ru/sveden/education/eduop/>

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle»

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) План практических занятий по дисциплине.

г) Методические указания по проведению практических занятий.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Ибрагимов И. М., Назаров Ю. Ф., Ковшов А. Н. Основы компьютерного моделирования наносистем. Москва Лань, 2010. 376 с.

– Полещук О.Х. Химические исследования методами расчета электронной структуры молекул : учебное пособие / О.Х. Полещук, Д.М. Кижнер. – Томск : Издательство ТПУ, 2006. – 146 с.

– Хельтзе Х.-Д. и др. Молекулярное моделирование: теория и практика: под ред. В. А. Палулина и Е. В. Радченко; пер. с англ. - М: Бином. Лаборатория знаний, 2009. -318 с.

б) дополнительная литература:

– Ермаков А. И. Квантовая механика и квантовая химия. В 2 ч. Часть 1. Квантовая механика : Учебник и практикум для вузов / Ермаков А. И.. - Москва : Юрайт, 2022. - 183 с - (Высшее образование) . URL: <https://urait.ru/bcode/491725>. URL: <https://urait.ru/book/cover/1C5DB5F7-EC6C-470E-A900-3CDE24C4E9D0>

– Соловьев М.Е. Компьютерная химия / – М.Е. Соловьев, М.М. Соловьев. – М. : ООО «СОЛОН-ПРЕСС», 2005. – 536 с.

– Ансельм А. И. Основы статистической физики и термодинамики: учеб. пособие. Москва Лань, 2007. – 423 с.

в) ресурсы сети Интернет:

– Научный журнал «Journal of Chemical information and modeling» – <https://pubs.acs.org/journal/jcisd8>

– Научный журнал «Molecular informatics» – <https://onlinelibrary.wiley.com/journal/18681751>

– Электронный учебный курс «Everything you need to get started in medical billing & coding» – <https://www.medicalbillingandcoding.org/what-is-mbac/>

– Научный журнал «Journal of Chemical information and modeling» – <https://pubs.acs.org/journal/jcisd8>

– Научный журнал «Molecular informatics» – <https://onlinelibrary.wiley.com/journal/18681751>

– Научный журнал «Journal of Cheminformatics» – <https://jcheminf.biomedcentral.com/>

– Виртуальная лаборатория «Virtual Computational Chemistry Laboratory» – <http://www.vcclab.org/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standard 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office Onenote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юрайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных:

– База данных «Protein Data Bank» – <http://www.rcsb.org>

– Спектральная база данных органических соединений «SDBS» – https://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

– База данных по рассчитанной квантово-химическими методами геометрии соединений «Computational Chemistry Comparison and Benchmark» – <https://cccbdb.nist.gov/geom1x.asp>

– База данных «Термические Константы Веществ» – <http://www.chem.msu.ru/cgi-bin/tkv.pl?show=welcome.html/welcome.html>

– База данных «ChEMBL» – <https://www.ebi.ac.uk/chembl>

– База данных «ChemSpider» – <http://www.chemspider.com>

– База данных «PubChem» – <http://pubchem.ncbi.nlm.nih.gov/>

– База данных «Reaxys» – <http://www.reaxys.com>

– База данных «ZINC» – <http://zinc.docking.org>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы и лабораторной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешанном формате («Актру»).

15. Информация о разработчиках

Фироз Неда, ассистент, Институт «Умные материалы и технологии» ТГУ.