

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

САЕ Институт «Умные материалы и технологии»

УТВЕРЖДАЮ:  
Директор

  
И.А. Курзина

« 05 » 11 2024 г.

Рабочая программа дисциплины

**Хемоинформатика**

по направлению подготовки

**19.03.01 Биотехнология**

Направленность (профиль) подготовки:  
**«Молекулярная инженерия»**

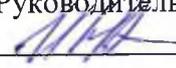
Форма обучения  
**Очная**

Квалификация  
**Бакалавр**

Год приема  
**2025**

СОГЛАСОВАНО:

Руководитель ОП

  
И.А. Курзина

Председатель УМК

  
Г.А. Воронова

## **1. Цель и планируемые результаты освоения дисциплины (модуля)**

Целью освоения дисциплины является формирование следующих компетенций:

–ОПК-2– Способен осуществлять поиск, хранение, обработку и анализ профессиональной информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий, включая проведение расчетов и моделирование, с учетом основных требований информационной безопасности;

–ПК-1– Способен к участию в проведении научно-исследовательской работы в своей профессиональной деятельности.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.3. Представляет результаты информационного поиска с использованием пакетов специализированных программ;

ИПК-1.2. Анализирует полученные данные и представляет результаты научных исследований по установленной форме.

## **2. Задачи освоения дисциплины**

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

– 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 Блоку 1 «Дисциплины (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, является обязательной для изучения.

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

Семестр 5, экзамен.

## **5. Входные требования для освоения дисциплины**

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Физическая химия», «Органическая химия».

## **6. Язык реализации**

Русский

## **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, тестирования и оценивания отчетов по выполненным лабораторным работам, и фиксируется в форме контрольной точки не менее одного раза в семестр.

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

Экзамен в пятом семестре проводится в форме устного опроса студентов по билетам. Продолжительность экзамена 1,5 часа.

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» <https://moodle.tsu.ru/course/view.php?id=34290>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине (<https://www.tsu.ru/sveden/education/eduop/>).

в) План семинарских занятий по дисциплине.

г) Методические указания по проведению лабораторных занятий.

### 12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Ибрагимов И. М., Основы компьютерного моделирования наносистем / Ибрагимов И. М., Ковшов А. Н., Назаров Ю. Ф. - Санкт-Петербург : Лань, 2022. - 384 с. URL: <https://e.lanbook.com/book/210257>

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

– Френкель Д. Принципы компьютерного моделирования молекулярных систем : от алгоритмов к приложениям / Даан Френкель, Беренд Смит ; пер. с англ. и науч. ред. В. А. Иванов, М. Р. Стукан ; [МГУ им. М. В. Ломоносова, Науч.-образоват. центр по нанотехнологиям]. - Москва : Научный мир, 2013. - 559 с.: ил. - ( Фундаментальные основы нанотехнологий : лучшие зарубежные учебники )

б) дополнительная литература:

– Степанов Н. Ф. Квантовая механика и квантовая химия : Учебник для химических факультетов университетов / Н. Ф. Степанов. - Москва : Мир [и др.], 2001. - 518, [2] с.: ил. - (Теоретические основы химии ) URL: <http://sun.tsu.ru/limit/2016/000142590/000142590.djvu>

– Соловьев М. Е. Компьютерная химия / Соловьев М. Е., Соловьев М. М.. - [Б. м. : б. и., 200-?]. - URL: <http://sun.tsu.ru/limit/2016/000344723/000344723.djvu>

– Ансельм А. И. Основы статистической физики и термодинамики : [учебное пособие для вузов по физическим и техническим направлениям и специальностям] / А. И. Ансельм. - Изд. 2-е, стер.. - Санкт-Петербург [и др.] : Лань, 2007. - 426 с.: ил.

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://sdb.sdb.aist.go.jp/sdb/cgi-bin/cre\\_index.cgi](https://sdb.sdb.aist.go.jp/sdb/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

Наименование оборудованных учебных кабинетов, объектов для проведения практических занятий, объектов физической культуры и спорта с перечнем основного оборудования	Адрес (местоположение) учебных кабинетов, объектов для проведения практических занятий, объектов физической культуры и спорта (с указанием площади и номера помещения в соответствии с документами бюро технической инвентаризации)
<p>Учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Аудитория № 115</p> <p>Оборудование: Графическая станция, процессор Intel i5, 16Гб оперативной памяти, монитор 24 дюйма</p> <p>Демонстрационный экран</p> <p>Мультимедиа-проектор</p> <p>Учебная мебель: рабочие места по количеству обучающихся (аудиторные столы, стулья); рабочее место преподавателя (стол, стул); аудиторная доска</p>	<p>634050, Томская область, г. Томск, пр-кт Ленина, 36, стр.7 (29 по паспорту БТИ) Площадь 40,9 м<sup>2</sup></p>
<p>Учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Аудитория № 103а</p> <p>Компьютерный класс: 13 компьютеров ASUS TUF B360-PLUS GAMING, Intel Core i7 8700, 16 ГБ, GeForce RTX 2070 8gb, 1 ТБ Жесткий диск; 240 Гб SSD, Блок питания CHIEFTEC GPC-700S 700 вт, Корпус Ginzzu A180 без БП,</p> <p>2. Монитор LG 24MK600M-B 23.8,</p> <p>3. Клавиатура Logitech K120, Мышь Logitech B100</p> <p>4. ПО, включающее:</p> <p>4.1. Windows 7</p> <p>4.2. Microsoft Office 2010</p> <p>4.3. Visual Studio 2019</p> <p>4.4. Visual Studio Code</p>	<p>634050, Томская область, г. Томск, пр-кт Ленина, 36, стр.7 (72 по паспорту БТИ) Площадь 43 м<sup>2</sup></p>
<p>Учебная аудитория для самостоятельной работы, индивидуальных консультаций. Аудитория № 121<sup>А</sup></p> <p>Учебная мебель: рабочие места по количеству обучающихся (аудиторные столы, стулья); рабочее место преподавателя (стол, стул)</p>	<p>634050, Томская область, г. Томск, пр-кт Ленина, 36, стр.7 (86 по паспорту БТИ) Площадь 23,8 м<sup>2</sup></p>

## **15. Информация о разработчиках**

Курзина Ирина Александровна, д-р ф-м. наук, зав. кафедрой ПСФимХ ХФ ТГУ