

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.

Оценочные материалы по дисциплине

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. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

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. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- тестирование
- отчеты по лабораторным работам

2.1 Лабораторная работа (ИОПК-2.3, ИПК-1.2):

С использованием базы ChEMBL отберите данные по активности соединений против циклооксигеназы-2 (COX-2). Отберите наиболее надежные данные. Проведите химическую и математическую чистку данных. Отберите 10% набора в качестве внешнего валидирующего набора случайным образом. С помощью программы ISIDA-MLR и различных типов фрагментных дескрипторов постройте несколько QSAR моделей (не менее 10). Для валидации модели используйте 5-кратную кросс-валидацию и внешний валидирующий набор. Отберите лучшую модель. Используйте эту модель для скрининга данных из базы ZINC. В качестве набора возьмите только «лекарствоподобные соединения» базы. Из числа соединений, показавших наибольшую активность ($IC_{50} < 10^{-7}$ моль/л), отберите 5%, имеющих наименьшую сумму расстояний от соединений, использованных для тренировки модели, в пространстве фрагментных дескрипторов.

Метод рекомендации по выполнению:

1. Подобрать по теме литературу и воспользоваться материалами, предоставленными на занятии преподавателем
2. Изучить вопросы практической работы в разных источниках
3. По необходимости проконсультироваться у преподавателя
4. Подготовиться к сдаче отчета по практической работе

Критерии оценивания:

Оценка «зачтено» выставляется в трех случаях:

5 баллов: если студент активно работает в течение всего практического занятия, дает полные ответы на вопросы преподавателя в соответствии с планом практического занятия и показывает при этом глубокое овладение лекционным материалом.

4 балла: если студент активно работает в течение практического занятия, вопросы освещены полно, изложения материала логическое, обоснованное фактами, со ссылками на соответствующие нормативные документы и литературные источники, освещение вопросов завершено выводами, студент обнаружил умение анализировать факты и события, а также выполнять учебные задания, но в ответах допущены неточности, некоторые незначительные ошибки.

3 балла: если студент в целом овладел сути вопросов по данной теме, обнаруживает знание лекционного материала и учебной литературы, пытается анализировать факты и события, делать выводы и решать задачи, но на занятии ведет себя пассивно, отвечает только по вызову преподавателя, дает неполные ответы на вопросы, допускает грубые ошибки при освещении теоретического материала.

Оценка «не зачтено» (2 и менее баллов) выставляется в случае, когда студент обнаружил несостоятельность осветить вопрос, вопросы освещены неправильно, бессистемно, с грубыми ошибками, отсутствуют понимания основной сути вопросов, выводы, обобщения, обнаружено неумение решать учебные задачи.

2.2 Тестовые задания (ИОПК-2.3, ИПК-1.2):

1. Какие из указанных SMILES соответствует молекуле аспирина?

- a. CC(=O)Oc1ccccc1C(O)=O
- b. c1c(C(O)=O)cc(OC(O)C)ccc1
- c. OC(=O)c(cccc1)c1OC(=O)C
- d. c1(C(=O)O)ccccc1OC(=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. OCCCCCCCCc1cc(C(=O)O)ccc1
- c. OCCc1ccc(C(C)=O)cc1
- d. OCCc1ccc(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? a - 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

Критерии оценивания тестовых заданий

Оценка «5» (отлично) – выставляется обучающемуся, допустившему до 10 % ошибок в тестовом задании.

Оценка «4» (хорошо) – выставляется обучающемуся, допустившему до 24 % ошибок в тестовом задании.

Оценка «3» (удовлетворительно) – выставляется обучающемуся, допустившему до 39 % ошибок в тестовом задании.

Оценка «2» (неудовлетворительно) – обучающийся допустил более 40 % ошибок в тестовом задании.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

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

Экзаменационный билет содержит 2 вопроса, проверяющий ИОПК-2.3 и ИПК-1.2. Ответ на вопрос дается в развернутой форме.

Примерные вопросы для экзамена:

1. Основные способы представления объектов в хемоинформатике;
2. Линейные представления;
3. Матричные представления;
4. Битовые представления;
5. Стандартные файлы в хемоинформатике;
6. Хемометрика и хемоинформатика;
7. Биоинформатика и хемоинформатика;
8. Дизайн библиотек данных;
9. Разбросанные и сфокусированные библиотеки;
10. Генерация структур;
11. RECAP;
12. Кластеризация молекул;
13. Иерархические и неиерархические подходы;
14. Отбор молекул без кластеризации;
15. Основные способы валидации моделей;
16. Классификационные модели;
17. Регрессионные модели;
18. Отбор дескрипторов;
19. Навигация в химическом пространстве как способ моделирования;
20. Методы машинного обучения.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценку «отлично» получают студенты, сдавшие все задания текущего контроля (получившие «зачтено» за каждый вид задания) и набравшие 8–9 баллов на экзамене, оценку «хорошо» получают студенты, сдавшие все задания текущего контроля (получившие «зачтено» за каждый вид задания) и набравшие 6–7 баллов на экзамене, оценку «удовлетворительно» получают студенты, полностью сдавшие все задания текущего контроля (получившие «зачтено» за каждый вид задания) и набравшие 4–5 баллов на экзамене, оценку «неудовлетворительно» получают студенты, сдавшие все задания текущего контроля (получившие «зачтено» за каждый вид задания) и набравшие

менее 4 баллов на экзамене. Студенты не сдавшие задания текущего контроля к экзамену не допускаются.

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Примеры тестовых вопросов (ИОПК-2.3, ИПК-1.2):

1. Хемоинформатика- это
 - а) это междисциплинарная область, объединяющая общую биологию, молекулярную биологию, кибернетику, генетику,
 - б) это мультидисциплинарное направление теоретической химии, находящееся на стыке химии, информатики, биологии, фармакологии, физики и математической статистики.**
 - в) наука о методах и процессах сбора, хранения, обработки, передачи, анализа и оценки химической информации
2. Дескрипторы
 - а) отражают факт наличия фрагмента в молекулярном графе (бинарные) или число вхождений фрагмента (целочисленные).**
 - б) фрагмент в молекулярном графе (бинарный)
 - в) трёхмерные геометрии и наиболее чувствительные к структурным изменениям
3. Возможно ли прогрессивное развитие химии без информатики?
 - а) да (обоснуйте ответ) _____
 - б) нет (обоснуйте ответ) _____
 - в) свой вариант _____
4. В какой из приведенных баз данных можно найти информацию, характеризующую прочность связывания данного химического соединения с различными белками?
 - а) CAS
 - б) PubChem**
 - в) ChEMBL
5. Постройте двуцепочечную молекулу ДНК: GGCCAACGAC. Изобразите ее в стилях wireframe, balls and sticks, van der waals. Для представления wireframe покажите водородные связи и кольца в данной структуре. Приведите все три изображения. (задание по Avogadro)

Ситуационная задача (ИОПК-2.3, ИПК-1.2):

Постановка задачи:

Фармацевтическая компания разрабатывает новое лекарство для лечения хронической боли. Исследователи синтезировали несколько молекул, которые демонстрируют перспективные результаты *in vitro*. Однако, прежде чем переходить к клиническим испытаниям, необходимо оценить фармакокинетические свойства этих молекул.

Ваша задача:

Используя методы хемоинформатики, предскажите следующие параметры для одной из перспективных молекул:

- * Липофильность (LogP): влияет на проникновение молекулы через клеточные мембраны и распределение в организме.
- * Молекулярная масса: влияет на скорость всасывания и выведения лекарства.
- * Количество доноров водородных связей: влияет на взаимодействие с белками-мишенями и растворимость в воде.

* Количество акцепторов водородных связей: влияет на взаимодействие с белками-мишенями и растворимость в воде.

Решение:

1. Загрузка молекулы: Загрузите молекулярную структуру в выбранное программное обеспечение для хемоинформатики.

2. Расчет параметров: Используйте соответствующие инструменты программного обеспечения для расчета параметров липофильности (LogP), молекулярной массы, количества доноров и акцепторов водородных связей.

3. Анализ результатов: Проанализируйте полученные значения. Например, высокая липофильность может указывать на хорошую проницаемость через клеточные мембраны, но также может увеличить риск накопления в организме. Низкая молекулярная масса может способствовать лучшему всасыванию, но также может привести к быстрому выведению.

4. Сопоставление с требованиями: Сравните предсказанные значения с оптимальными фармакокинетическими параметрами для лекарственных средств данного класса. Например, для препаратов, предназначенных для перорального приема, часто требуется определенный диапазон LogP, молекулярной массы и т.д.

Ответ: В результате проведенного анализа получены следующие значения:

* LogP: [Вставьте полученное значение]

* Молекулярная масса: [Вставьте полученное значение]

* Количество доноров водородных связей: [Вставьте полученное значение]

* Количество акцепторов водородных связей: [Вставьте полученное значение]

Выводы: На основе полученных значений можно сделать выводы о потенциальных фармакокинетических свойствах исследуемой молекулы. Например, если LogP слишком высокое, может потребоваться изменение структуры молекулы для повышения ее водорастворимости.

Информация о разработчиках

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