

Вопросы государственного экзамена программы магистратуры “Применение машинного обучения в биологии” (направление подготовки 06.04.01 «Биология»)

1. Белки: строение и функции в клетке. Свойства аминокислот, однобуквенное и трехбуквенное обозначения. Аминокислоты как кислоты и основания. Пространственная структура полипептидов. Нековалентные и дисульфидные связи. Элементы вторичной структуры белка (альфа-спирали, бета-тяжи). Фолдинг и денатурация белков. Глобулярные и мембранные белки. Примеры транспортных белков. Моторные белки. Ферменты.
2. Нуклеотиды и нуклеиновые кислоты. Структура, номенклатура и свойства нуклеотидов. Фосфодиэфирная связь, понятие комплементарности нуклеотидных цепей. Структура ДНК и РНК. Метилирование азотистых оснований. Функции свободных нуклеотидов в клетке: АТФ, редокс-кофакторы (НАДН, НАДФН, FAD), цАМФ.
3. Липиды. Жирные кислоты: насыщенные и ненасыщенные. Триацилглицериды как энергетический резерв клетки. Структурные липиды мембран. Сигнальные липиды: фосфатидилинозитол и эйкозаноиды, стероидные гормоны. Хиноны как переносчики электронов. Строение биологических мембран. Липидный бислой, мембранные белки. Транспорт веществ через мембраны: простая и облегченная диффузия, активный транспорт. Симпорт и антипорт. Первичный и вторичный активный транспорт. Аквапорины, натрий-калиевая АТФаза, ABC-переносчики. Ионифоры.
4. Скорость химической реакции, факторы, от которых зависит скорость реакции. Тепловой эффект реакции. Энергия активации. Закон Гесса.
5. Окислительно-восстановительные реакции, окислительно-восстановительный потенциал. Уравнение Нернста.
6. Понятия энтропии, энтальпии. Энергия Гиббса. Законы термодинамики.
7. Систематика и биоразнообразие живого мира. Клеточная теория. Внутриклеточные органеллы эукариотической клетки. Строение и функции ядра клетки. Ядрышко и рибосомы. ЭПР, комплекс Гольджи, лизосомы, вакуоли и пероксисомы – элементы вакуолярной системы внутриклеточного транспорта. Митохондрии и пластиды: строение, функции, гипотеза симбиогенеза. Цитоскелет. Центриоли. Органеллы движения – реснички, жгутики и псевдоподы.
8. Жизненный цикл клетки. Основные процессы жизнедеятельности в пресинтетический, синтетический и постсинтетический период. Строение хромосом и пloidность клетки. Понятие о гомологичных хромосомах. Способы деления клетки. Бинарное деление прокариотической клетки. Митотическое деление клеток. Мейоз. Конъюгация и кроссинговер.
9. Основные виды мутаций: генные (транзиции и трансверсии, синонимичные и несинонимичные мутации), хромосомные, геномные. Основные причины мутаций каждого вида. Примеры заболеваний, вызванных мутациями (генные - муковисцидоз, хорей Хантингтона, хромосомные – синдром Лежена, геномные – синдром Дауна, Клайнфельтера, Шерешевского-Тернера). Диагностика носительства мутаций.

10. Гомологичная и сайт-специфическая рекомбинация. Кроссинговер.
11. Транскрипция, ее основные принципы, промоторы, ферменты. Опероны прокариот, регуляция их работы. Особенности транскрипции эукариот. Промоторы, энхансеры, транскрипционные факторы. Процессинг РНК. Сплайсинг, его основные механизмы, участие малых ядерных РНК. Альтернативный сплайсинг, его биологическое значение.
12. Метилирование ДНК как способ регуляции транскрипции. 5mC и 5hmC основные продукты метилирования-деметилирования у эукариот. Обратимость метилирования, стохастическая модель метилирования. Модификации гистонов. Гистоновый код. Значение эпигенетической регуляции экспрессии генов.
13. Регуляторные РНК. РНК-переключатели прокариот. МикроРНК. Сайленсинг, общий механизм. Si- и sh-РНК, их использование для нокдауна генов. Длинные некодирующие РНК. Инактивация X-хромосомы млекопитающих РНК Xist.
14. Трансляция. Структура зрелой мРНК эукариот, функции кэпа и поли А-хвоста. Структура и состав рибосом, 70S и 80S рибосомы, полисомы. Структура и функциональные участки тРНК. Аминоацилирование. Основные стадии трансляции (инициация, элонгация, терминация). Различия в механизмах инициации у про- и эукариот: последовательность Шайна-Дальгарно, роль кэпа и инициаторных факторов. IRES-элементы. Котрансляционное сворачивание полипептидной цепи, шапероны. Посттрансляционные модификации белков, их биологическое значение.
15. Деградация белков: регуляция их активности и времени жизни. Посттрансляционный протеолитический процессинг (активация белков-предшественников, отщепление сигнальных последовательностей). Автопротеолиз. Лизосомальная деградация белков – аутофагия. Протеасомная деградация белков: структура протеасомы, механизм ее работы. Роль убиквитинилирования в протеасомной деградаци белка. Убиквитин-независимая деградация белков в протеасомах. Примеры заболеваний, связанных с нарушением протеасомной деградаци.
16. Выделение, очистка, количественная и качественная оценка нуклеиновых кислот. Методы детекции и гибридизации. Полимеразная цепная реакция и ее основные варианты, в том числе RT-ПЦР. Рестрикционный анализ. Секвенирование ДНК. Библиотеки нуклеиновых кислот. Возможности и ограничения основных методов молекулярной генетики.
17. Генная инженерия как направленное изменение генетического материала и способ изучения функции генов. Основные принципы клонирования ДНК. Векторы и их типы. Получение рекомбинантных молекул ДНК. Введение генетических конструкций в клетки. Подходы к изменению экспрессии и функции генов: нокаут, нокдаун, РНК-интерференция, редактирование генома. Связь генотипа и фенотипа. Подходы к экспериментальной проверке того, что внесенное генетическое изменение действительно произошло и привело к ожидаемому биологическому эффекту.
18. Вирусы как биологические системы: строение, геномы, классификация и репродуктивные стратегии. Место вирусов в системе живого. Строение вирусов, капсиды и оболочки. Типы вирусных геномов, сегментированные и многокомпонентные геномы. Проникновение вирусов в клетку. Основные схемы репликации ДНК- и РНК-вирусов и стратегии экспрессии вирусных геномов.
19. Вирусы как патогены и как инструменты биологии и биотехнологии. Бактериофаги, вирусы растений, вирусы животных и человека. Литический и лизогенный циклы. Патогенез вирусных инфекций, механизмы повреждения клеток, онкогенные вирусы, зоонозы и новые вирусные

инфекции. Использование вирусов и бактериофагов в биоинженерии для трансформации организмов.

20. Прокариоты: разнообразие, строение, эволюция и метаболизм. Разнообразие микроорганизмов, место бактерий и архей в живом мире. Строение прокариотической клетки и особенности организации генома. Горизонтальный перенос генов. Разнообразие вариантов метаболизма прокариот, их роль в биосфере и биогеохимических процессах. Бактериальные сообщества. Значение микроорганизмов в биотехнологии и медицине.
21. Патогенные бактерии, антибиотики и диагностика бактериальных инфекций. Патогенные и условно-патогенные бактерии, факторы патогенности, внутриклеточный паразитизм. Антибиотики, их мишени и механизмы действия. Природа и распространение антибиотикорезистентности, включая мультирезистентность. Подходы к выявлению возбудителя и его чувствительности к терапии, включая биоинформатические методы и методы искусственного интеллекта.
22. Врождённый иммунитет: PRR, PAMP, DAMP. Ключевые клеточные игроки: нейтрофилы, макрофаги, NK-клетки. Дендритные клетки и MHC-презентация. Адаптивный иммунитет: T-клетки, тимусная селекция. B-клетки, изотипы антител, гипермутации. Иммунная память. Примеры «пограничных» популяции иммунных клеток, не укладывающихся в дихотомию врожденный/адаптивный иммунитет.
23. Строение TCR и BCR, разнообразие иммунного репертуара, VDJ-рекомбинация как основа огромного разнообразия антител. Роль CDR3 в разнообразии и распознавании антигена. AIRR-анализ: определение и биологические задачи, использование UMI. Разнообразие рецепторов адаптивного иммунитета, метрики оценки разнообразия (chao1, нормированная энтропия Шеннона-Винера). Публичные клонотипы и конвергентный иммунный ответ. Вычислительные методы предсказания специфичности TCR. Анализ BCR: SHM и деревья клонотипов.
24. Биология опухолей. Общие свойства опухолей, отличия доброкачественных и злокачественных опухолей. Типы злокачественных опухолей. Причины возникновения опухолей. Онкогены и гены-онкосупрессоры. Локомоторный фенотип как основа агрессивных свойств опухолей.
25. Метастазирование опухолей, его типы и механизмы. Роль иммунной системы в опухолевой прогрессии. Экспериментальное воспроизведение опухолей. Современные принципы и подходы к диагностике и терапии опухолей: таргетная и персонализированная терапия.
26. Организация генома человека. Характеристики белок-кодирующих генов. Механизмы и частота возникновения мутаций в геноме человека. Основные процессы в популяционной генетике, определяющие динамику аллелей. Варианты генома в индивидуумах и популяциях. Различные типы вариантов генома, их функциональный эффект и клиническая значимость. Неравновесие по сцеплению.
27. Моногенные заболевания: примеры, поиск каузальных генов и вариантов. Различные виды наследования: аутосомно-доминантное и др. Пенетрантность. Диагностика редких заболеваний с помощью высокопроизводительного секвенирования. Критерии для клинической интерпретации результатов секвенирования.
-
28. Асимптотический анализ, определение O-сложности. Алгоритмы сортировки массива и их асимптотические сложности. Теорема о нижней границе сортировки сравнениями.
29. Понятие графа, виды графов. Алгоритмы на графах. Обходы графа. Задача поиска кратчайшего пути в графе. Динамическое программирование, алгоритм Дейкстры, алгоритм Беллмана.

30. Структуры данных в алгоритмах. Массив, список, дерево, куча, хэш-таблица. Абстрактные структуры: стек, очередь. Основные операции на структурах и их асимптотические сложности.
31. Производная функции. Матрица Якоби, градиент. Градиентный спуск. Эвристики градиентного спуска: метод импульса, метод Нестерова, AdaGrad, RMSProp, Adam.
32. Векторное пространство, определение. Размерность пространства. Понятие линейной оболочки. Понятие ранга матрицы, связь ранга матрицы с ядром и образом.
33. Собственные значения и собственные векторы. Спектральная теорема. Спектральное разложение матрицы. Связь с сингулярным разложением.
34. Комбинаторика: перестановки, сочетания, размещения с повторениями и без. Рекуррентные формулы, число композиций и разбиений.
35. Вероятностное пространство, событие. Определение вероятности. Вероятность объединения и пересечения событий. Независимые и несовместные события. Условная вероятность, теорема Байеса. Априорная и апостериорная вероятности. Метод максимального правдоподобия.
36. Случайная величина. Дискретные и непрерывные случайные величины. Распространенные дискретные распределения: Бернулли, биномиальное, Пуассона, геометрическое, гипергеометрическое, отрицательное биномиальное. Распространенные непрерывные распределения: равномерное, экспоненциальное, нормальное, Стюдента, хи-квадрат.
37. Мат. ожидание и дисперсия случайной величины. Свойства среднего и дисперсии. Ковариация и корреляция, отношение корреляции к причинно-следственной связи.
38. Устройство статистического теста. H_0 , H_1 . Определение p-value, уровень значимости, ошибки 1 и 2 рода. Параметрические и непараметрические тесты. Статистические тесты для выявления взаимосвязи категориальных и/или числовых переменных.
39. Проблема множественного тестирования. FWR и FDR.
40. Типы баз данных NCBI. Геномные данные. NCBI GenBank и NCBI RefSeq. Принцип сборки генома, риды, контиги, скаффолды, N50, L50.
41. Алгоритм BLAST: подвиды алгоритма, общий метод работы. Параметры находок алгоритма BLAST: score, e-value, identity, gaps, strand. Определение E-value.
42. Выравнивание как отражение эволюции последовательностей. Алгоритмы выравниваний: Нидлмана - Вунша, Смита - Уотермана. Матрица весов в алгоритме выравнивания: матрицы PAM и BLOSUM.
43. Транскрипционные факторы. Мотивы, вхождения мотивов, PWM. Лого-диаграммы. Поиск мотивов связывания ТФ. Алгоритм MEME.
44. Технология секвенирования. Секвенирование по Сэнгеру. Секвенирование нового поколения: платформы Illumina, MGI, Oxford Nanopore, PacBio. Методы контроля качества результатов секвенирования: удаление адаптеров, тримминг. RNA-seq, ChIP-seq, Hi-C. Картирование прочтений, диф.экспрессия.
45. Методы аннотации геномов. Предсказание генов ab initio, по гомологии и известным данным.
46. Модели последовательностей: Бернуллиевская, Марковские. Скрытые Марковские модели. Генерация и декодирование.
47. Нековалентные взаимодействия в структурах белков. Дисперсионные взаимодействия. Ван-дер-Ваальсовы взаимодействия. Водородная связь, ее природа и геометрия. Гидрофобный эффект. π -Стекинг, π -Катионное взаимодействие. Солевые мостики. Галогеновое связывание. Нековалентные взаимодействия боковых радикалов аминокислот.
48. Вторичная структура белка, типы, свойства, паттерны водородных связей. Алгоритмы DSSP, Stride: применение к разметке вторичной структуры. Функциональные мотивы как пространственные паттерны атомов (на примере каталитических сайтов и P-loop мотива).

49. Третичная структура белка и понятие домена: эволюционные, функциональные, структурные и тривиальные домены. Классификация доменов и укладок (SCOP и CATH). Алгоритмы выделения структурных доменов: DOMAK, DETECTIVE.
50. Физико-химические методы определения структуры биомакромолекул: PCA, КриоЭМ, ЯМР. Основные принципы работы. Применимость, сильные и слабые стороны методов. Основные источники ошибок в получаемых структурах. Предсказание структуры белков по последовательности. AlphaFold, AlphaFold2, AlphaFold3: эволюция метода. Метрики качества моделей. Ошибки в предсказанных моделях.
51. Связь квантовой химии, молекулярной механики и молекулярной динамики в моделировании молекулярных систем. Главные особенности молекулярно механических силовых полей. Молекулярная динамика. Потенциальная энергия силового поля как основа уравнений движения. Этапы подготовки системы для молекулярной динамики. Исследование потенциальной поверхности энергии методами расширенного сэмплинга. Коллективные переменные как описание релевантных координат на поверхности свободной энергии.
52. Процесс создания лекарственных молекул, основные этапы. Проблематика выбора молекулярной мишени. Поиск соединений-хитов. Первичный *in vitro* скрининг, подтверждающие эксперименты. Понятие об IC50, применимость IC50 в зависимости от типа мишени. Патенты, организационные мероприятия по интеллектуальной защите. Hit to lead оптимизация. Основные параметры для оптимизации. Понятие об ADME: основные ADME-параметры, требующие оптимизации. Lead optimization. Клинические исследования.
53. На что воздействуют лекарства-малые молекулы. Основные классы молекулярных мишеней. Распределение мишеней по клиническим показаниям. Анти-мишени. Ортостерические и аллостерические лекарства: принцип действия, преимущества и недостатки. Агонисты, антагонисты, обратные агонисты. GPCR-рецепторы как ключевой класс мишеней, особенности разработки лигандов для GPCR. Подходы к дизайну при отсутствии структурной информации о мишени. Протеиновые киназы как класс ферментных мишеней. Типы киназных ингибиторов. Понятие фармакофора. Классический киназный фармакофор. Фармакофор в ЦНС-химии как отражение свойств лигандов GPCR и других рецепторов.
54. ADME. Фармакокинетика и фармакодинамика. Введение ксенобиотиков. Всасывание и проникающая способность: активный и пассивный транспорт. Роль транспортеров, примеры (P-gp). Связывание с белками плазмы. Распределение. Объем распределения. Метаболизм и его фазы, основные реакции. Стратегии по оптимизации метаболизма. Выведение ксенобиотиков. Фармакокинетические модели для внутривенного и перорального введения. Определение пероральной биодоступности, клиренс. DDI (drug-drug interactions): CYP P450 и P-gp, субстраты, ингибиторы, активаторы (индукторы).
55. Пептидомиметики. Преимущества, сфера использования, лекарственные мишени пептидомиметиков. Главные структурные аспекты при создании протеазных ингибиторов, терапевтические области применения. Понятие о макроциклических ингибиторах протеаз.
56. Машинное обучение. Supervised обучение. Разбиение данных на тренировочную и тестовую выборки. Валидация модели машинного обучения. Недообучение и переобучение модели. Дилемма смещения-дисперсии.
57. Задача регрессии. Линейная регрессия, аналитический вывод линейной регрессии методом наименьших квадратов. Регуляризация, виды регуляризации. Эффект регуляризации на веса модели. Задача классификации, логистическая регрессия. Метрики качества классификации.
58. Unsupervised обучение. Задача кластеризации, понятие кластера. Виды и принципы работы алгоритмов кластеризации. Задача снижения размерности. Проклятие размерности. Гипотеза о

многообразии (manifold hypothesis). Методы снижения размерности: анализ главных компонент (PCA), t-SNE, UMAP.

59. Модели, основанные на деревьях решений. Ансамбли моделей: бэггинг и бустинг. Случайный лес. AdaBoost. Градиентный бустинг. Особенности моделей XGBoost, LightGBM, CatBoost.
60. Нейронные сети. Пайплайн обучения нейронной сети. Многослойный перцептрон. Функции активации в MLP: логистическая функция, tanh, ReLU.
61. Сверточные нейронные сети. Слои свертки (convolution), снижения (pooling) и повышения разрешения (upsampling). Параметры сверточных слоев. Понятие карты признаков сверточной сети. Рецептивное поле сверточной нейросети.
62. Рекуррентные нейронные сети (RNN). Архитектуры основных рекуррентных нейронных сетей: GRU, LSTM. Механизм внимания. Архитектура трансформера.
63. Понятия эмбединга и скрытого (латентного) пространства. Автоэнкодер, концепция и архитектура модели. Вариационный автоэнкодер. Генерация объектов с помощью VAE.
64. Генеративное моделирование. Состязательные атаки, генеративно-состязательные нейронные сети. Диффузионные нейронные сети (DDPM): постановка задачи и принцип работы. Архитектура Stable Diffusion. Принцип работы FrameDiff в RFDiffusion.