

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ
«НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ БИОМЕДИЦИНСКОЙ
ХИМИИ ИМЕНИ В.Н. ОРЕХОВИЧА» (ИБМХ)**

«ПРИНЯТО»

На заседании Ученого совета ИБМХ

Протокол № 4 от «21» апреля 2022 г.



«УТВЕРЖДАЮ»

Директор ИБМХ

Е.А. Пономаренко

«22» апреля 2022 г.

ПРОГРАММА

**для сдачи кандидатского минимума
по специальности 1.5.8 - математическая биология, биоинформатика
на соискание ученой степени кандидата биологических наук**

Москва 2022 г.

Паспорт научной специальности 1.5.8. «Математическая биология, биоинформатика»

Область науки:

1. Естественные науки

Группа научных специальностей:

1.5. Биологические науки

Наименование отрасли науки, по которой присуждаются ученые степени:

Физико-математические

Биологические

Медицинские

Шифр научной специальности:

1.5.8. Математическая биология, биоинформатика

Направления исследований:

1. Математическое и компьютерное моделирование живых систем: биомолекул, ферментативных реакций, метаболических и сигнальных путей, субклеточных структур, клеток, тканей, органов, систем органов, организмов, популяций, биоценозов.
2. Компьютерная системная биология (геномика, транскриптомика, протеомика, метаболомика, другие омиксные исследования).
3. Математическое и компьютерное моделирование структурно-функциональных взаимоотношений отдельных биомолекул и их взаимодействий в клетке (интерактомика).
4. Математическое и компьютерное моделирование биологического действия ксенобиотиков. Компьютерное конструирование лекарств. Анализ взаимосвязей «структура-активность». Компьютерная фармакология и токсикология.
5. Идентификация потенциальных биомаркеров с целью диагностики заболеваний и перспективных молекулярных мишеней новых лекарств.
6. Компьютерное конструирование иммуногенных конструкций с целью создания новых вакцин. Белковая инженерия. Конструирование антител.
7. Компьютерное конструирование микроорганизмов и растений с требуемыми свойствами.
8. Математическое и компьютерное моделирование эволюционных процессов в живой природе.
9. Математическое и компьютерное моделирование экологических систем.
10. Разработка новых вычислительных технологий на основе результатов исследований живых систем; развитие бионических подходов.
11. Организация, ведение и использование специализированных мультидисциплинарных банков данных и баз знаний по биологии и медицине, в т.ч. банков междисциплинарных данных.
12. Разработка и применение новых вычислительных алгоритмов для анализа экспериментальных данных в биологии и медицине.
13. Компьютерное распознавание, анализ и синтез изображений в биологических и медицинских исследованиях.

14. Математические модели, численные методы, алгоритмы и программные средства применительно к процессам получения, накопления, обработки и систематизации биологических и медицинских данных и знаний.
15. Математический и компьютерный анализ биомедицинских текстов, извлечение информации о биологических объектах и их взаимосвязях.
16. Разработка и применение методов машинного обучения и искусственного интеллекта для анализа и прогнозирования свойств биологических объектов на основе анализа больших биомедицинских данных.

Смежные специальности (в т.ч. в рамках группы научной специальности):

- 1.5.2. Биофизика
- 1.5.3. Молекулярная биология
- 1.5.4. Биохимия
- 1.5.6. Биотехнология
- 1.5.7. Генетика
- 1.5.22. Клеточная биология

Базовые понятия современной молекулярной биомедицины

1. Основная догма молекулярной биологии. Матричный принцип. Процессы репликации, транскрипции, трансляции. Генетический код.
2. Структурная организация генов и геномов. Экзон-интронная структура. Хромосомы.
3. Транскрипция и её регуляция. Транскрипционные факторы. Типы регуляторных районов транскрипции. Структура и функция промотора.
4. Структурно-функциональные характеристики белков и нуклеиновых кислот.
5. Структура и функция РНК. Экспериментальные методы определения структуры РНК.
6. Трансляция РНК. Регуляция трансляции.
7. Структура и функция белков. Выравнивание белковых структур. Распознавание функциональных сайтов и мотивов в белках.
8. Функциональная геномика. Понятие экспрессии генов. Экспериментальные методы анализа экспрессии.
9. Посттрансляционные модификации в белках. Структурно-функциональное значение. Методы экспериментального определения и компьютерного предсказания.
10. OMICs науки XXI века. Геномика, транскриптомика, протеомика, метаболомика.
11. Системная биология. Генные и метаболические сети, регуляторно-сигнальные пути. Экспериментальное изучение и компьютерное моделирование.
12. Биомаркеры и фармакологические мишени. Методы определения и валидации.
13. Путь от гена к лекарству *in silico*. Компьютерное конструирование лекарств.
14. Персонализированная медицина. Понятие об исследованиях ассоциаций (GWAS) Современное состояние и перспективы развития.
15. Основные экспериментальные методы молекулярной биологии. Клонирование, ПЦР, секвенирование по Сенгеру, Массовое секвенирование (NGS) и связанные с ним подходы. Масс-спектрометрия.
16. Система CRISPR/Cas9. Возможности для редактирования генома и использования в терапевтических целях.
17. Филогения и эволюционные деревья. Подходы к изучению филогенеза, видового разнообразия и эволюционных взаимоотношений на основе геномных и протеомных исследований.
18. Микробиом человека в норме и при патологиях.
19. Метагеномика. Основные экспериментальные методы. Общий алгоритм анализа метагеномных данных. Практическое применение результатов метагеномных исследований.
20. Генетика популяций и эволюция.
21. Связь между особенностями индивидуального генома человека и заболеваниями. Классификация заболеваний на основе такой связи. Связь между частотой вариаций и величиной патогенного эффекта.
22. Клиническая классификация генетических вариантов. Количество вариаций в индивидуальном геноме. Аннотация вариаций. Основные базы данных и компьютерные программы для поиска и анализа клинически значимых SNP. Сравнение полногеномных и полноэкзомных исследований в клинике.

Информационные ресурсы и алгоритмы биоинформатики

1. Теория вероятностей и статистика. Случайные величины. Статистические критерии, Понятие о множественном тестировании. Байесов подход, понятие об априорной и апостериорной вероятности.

2. Методы машинного обучения. Решающие деревья, случайный лес, метод поддерживающих векторов (SVM), регрессионный анализ, кластерный анализ, факторный анализ. Оценка качества обучения. Проблема переобученных моделей.
3. Машинное обучение в биоинформатике, его основные составляющие.
4. Наивный Байесовский классификатор, инвариантная точность прогноза.
5. Классификация на основе сходства, мера Жаккара.
6. Метод опорных векторов, критерии качества классификации.
7. Метод наименьших квадратов, критерии качества зависимостей.
8. Нелинейная регрессия, логистическая регрессия.
9. Методы регуляризации в задаче восстановления зависимостей по эмпирическим данным, самосогласованная регрессия.
10. Методы оценки результатов машинного обучения.
11. Искусственные нейронные сети, их возможности и ограничения.
12. Методы кластеризации.
13. Основные понятия о базах данных. Структура и содержание базы данных. Таблицы, поля, запросы.
14. Основные информационные ресурсы и базы данных по молекулярной биологии. Архивные, курируемые и производные базы данных.
15. Понятия Data mining и Text mining.
16. Понятия «коллекция» и «корпус» текстов. Принципы обработки и представления текстов для распознавания наименований объектов и выявления взаимосвязей между объектами. Основные подходы, алгоритмы, инструменты для распознавания наименований объектов в текстах и выявления взаимосвязей между объектами в текстах.
17. Основные информационные ресурсы по нуклеотидным последовательностям, принципы пополнения данных. Типы данных по нуклеотидным последовательностям: риды, сборки, контиги, скаффолды.
18. UniProtKb, принципы пополнения данных, структура данных, SWISS-PROT и TrEMBL.
19. Информационные ресурсы по функциональной характеристике генов и белков: Gene Ontology, KEGG, OMIM, Protein Atlas.
20. Парное выравнивание аминокислотных последовательностей на основе динамического программирования. Эвристический алгоритм BLAST.
21. Иерархическое множественное выравнивание аминокислотных последовательностей.
22. Идентификация доменов в аминокислотных последовательностях с использованием профилей. Информационные ресурсы по доменам и белковым семействам.
23. Основные подходы к построению филогенетических деревьев. Ортологи и паралоги.
24. Функциональная аннотация белков по последовательностям. Предсказание аминокислотных позиций, определяющих функциональную специфичность.
25. Методы секвенирования нового поколения (NGS). Формат данных. Контроль качества данных. Основной алгоритм анализа NGS данных.
26. Анализ профилей генной экспрессии. Примеры и цели транскриптомных исследований. Основные базы данных. Экспериментальные технологии получения транскриптомных данных: микрочипы и RNASeq.
27. Что понимают под межмолекулярными взаимодействиями. Экспериментальные и биоинформатические подходы к выявлению белок-белковых взаимодействий. Базы данных по интерактомике. Виды и свойства биологических сетей. Способы анализа сетей, топология.

28. Основные свойства молекулярных сетей. Понятия «степень вершины», «хаб», «кратчайший путь», «связность», «центральность», «мотив». Понятие о модулях в сети и их связи друг с другом.
29. Предсказание генов (белков), связанных с заболеваниями, на основе анализа молекулярных сетей.

Системная фармакология и молекулярное моделирование

1. Экспериментальные и теоретические подходы к поиску и разработке лекарств.
2. Основные информационные ресурсы по биологически активным соединениям.
3. Три основных способа получения данных из базы данных ChEMBL и их особенности. Основные причины наличия противоречий в данных по оценке биологической активности химических соединений в базах данных.
4. Поиск и приоритизация фармакологических мишеней. Понятие и оценки "Druggability".
5. Оценка фармакологических мишеней на основе анализа топологии сигнальных регуляторных сетей. Понятие о «мастер-регуляторах».
6. Оценка фармакологических мишеней на основе дискретного моделирования клеточного сигналинга. Логические сети и сети Петри: основные понятия и этапы анализа.
7. Поиск и оптимизация базовых структур новых лекарств. Оценки "Druglikeness" и "Leadlikeness".
8. Методы конструирования лекарств на основе структуры лигандов.
9. Прогнозирование метаболизма ксенобиотиков *in silico*. Определение понятия «метаболизм ксенобиотиков». Фазы метаболизма, основные ферментные системы и реакции биотрансформации.
10. Понятие «метаболическая стабильность». Влияние различных факторов на скорость метаболизма. Экспериментальные методы анализа метаболической стабильности и структур метаболитов.
11. Базы данных и компьютерные методы прогноза метаболизма ксенобиотиков.
12. Оценка безопасности базовых структур новых лекарств *in silico*.
13. Плейотропное действие лекарственных соединений: положительные и отрицательные аспекты.
14. Методы компьютерного предсказания межлекарственного взаимодействия (МЛВ): (Q)SAR, PBPK, молекулярное моделирование, машинное обучение. Базы данных и онлайн сервисы для оценки и прогноза МЛВ.
15. Структура белка. Методы получения трехмерной структуры белка. Визуализация белковых структур и комплексов. Банк данных пространственных структур белков PDB.
16. Инструменты для интерактивной визуализации белковых структур. Выявление сходных 3-мерных структур белков. Визуализация белковых структур и изучение свойств белковых молекул при помощи программы PyMol.
17. Методы предсказания белковых структур по последовательностям аминокислот. Моделирование трехмерной структуры белка методом гомологического моделирования AlphaFold2.
18. 3D методы в конструировании лекарств на основе структуры лигандов. Фармакофорная модель. 3D-QSAR.
19. Анализ белок-лигандного взаимодействия методами молекулярного моделирования. Модели комплексов, силы, участвующие в связывании лигандов, термодинамика и кинетика процесса.
20. Поиск и конструирование новых лигандов. Молекулярный докинг, конструирование *de novo*, оптимизация лигандов.
21. Моделирование молекулярной динамики: возможности и ограничения.

Основная литература

1. Введение в информационную биологию и биоинформатику. Под ред. Н.А. Колчанова, О.В. Вишневого, Д.П. Фурман. Том 2. Компьютерная протеомика. Новосибирск: Новосибирский гос. ун-т, 2012. 252 с.
2. Введение в информационную биологию и биоинформатику. Под ред. Н.А. Колчанова, О.В. Вишневого, Д.П. Фурман. Том 3. Теория генных сетей. Картирование генов, контролируемых сложные признаки человека. Новосибирск: Новосибирский гос. ун-т, 2015. 298 с.
3. Введение в информационную биологию и биоинформатику. Под ред. Н.А. Колчанова, О.В. Вишневого, Д.П. Фурман. Том 4. Математическое моделирование и методы биоинформатики в биологии развития. Компьютерная эволюционная биология Новосибирск: Новосибирский гос. ун-т, 2012. 336 с.
4. Маджидов Т.И. Введение в хемоинформатику: Компьютерное представление химических структур: учеб. пособие / Т.И. Маджидов, И.И. Баскин, И.С. Антипин, А.А. Варнек. – Казань: Казан. ун-т, 2013. – 174 с.
5. Маджидов, Т.И. Введение в хемоинформатику: учеб. пособие. Ч. 2. Химические базы данных / Т.И. Маджидов, И.И. Баскин, А.А. Варнек. - Казань: Изд-во Казан. ун-та, 2015. - 188 с.
6. Баскин, И.И. Введение в хемоинформатику: учеб. пособие. Ч. 3. Моделирование "структура-свойство" / И.И. Баскин, Т.И. Маджидов, А.А. Варнек. - Казань: Изд-во Казан. ун-та, 2015. - 304 с.
7. Введение в хемоинформатику: учеб. пособие. Ч. 4. Методы машинного обучения / И.И. Баскин, Т.И. Маджидов, А.А. Варнек. – Казань: Изд-во Казан. ун-та, 2016. – 330 с.
8. Введение в хемоинформатику: учеб. пособие. Ч. 5. Информатика химических реакций / И.И. Баскин, Т.И. Маджидов, А.А. Варнек. – Казань: Изд-во Казан. ун-та, 2017. – 244 с.
9. Введение в хемоинформатику: учеб. пособие. Ч. 6. Химическое пространство и виртуальный скрининг / Т.И. Маджидов, И.И. Баскин, А.А. Варнек. – Казань: Изд-во Казан. ун-та, 2019. – 240 с.
10. Поройков В.В. В кн.: Доклиническое изучение лекарственных средств (промышленная фармация): учебник для студ. высш. учебн. заведений. Под ред. А.Л. Хохлова, Н.В. Пятигорской. М.: Из-во ООО «ГРУППА РЕМЕДИУМ», 2021. С. 28-80.
11. Раевский О.А. Моделирование соотношений «структура-свойство». М.: Из-во «КДУ», 2015. 288 с.
12. Финкельштейн А.В., Птицын О.Б. Физика белка. М.: Книжный дом «Университет», 2002.
13. Каменская М.А. Информационная биология. М.: Издательский центр «Академия», 2006.
14. Хельтс Х.-Д., Зиппл В., Роньян Д., Фолькерс Г. Молекулярное моделирование. Теория и практика. М.: Издательство «Бином», 2010.
15. Жимулёв И.Ф. Общая и молекулярная генетика. Учебное пособие. Новосибирск. НГУ. 2003.
16. Ризниченко Г. Ю. Лекции по математическим моделям в биологии. Издание 2-е, исправленное и дополненное. Изд-во РХД, М–Ижевск, 2011 г. 560 с.
17. Шайтан К.В., Сарайкин С.С. Молекулярная динамика. М.: МГУ им. М.В. Ломоносова, 1999.
18. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. М.: Медицина, 1975.

19. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. М.: Финансы и статистика, 1989.
20. Химмельблау Д. Прикладное нелинейное программирование, М.: Мир, 1975.
21. Дурбин Р., Эдди Ш., Крэг А., Митчисон Г. Анализ биологических последовательностей. Издательство: Регулярная и хаотическая динамика, Институт компьютерных исследований, 2006
22. Lenghauer T. Bioinformatics. From Genomes to Drugs. Weinheim: Wiley-VCH, 2002.

Дополнительная литература

1. Свердлов Е.Д. Взгляд на жизнь через окно генома. М.: Наука. Т.1: Очерки структурной молекулярной генетики. - 2009. - 525 с.
2. Свердлов Е.Д. Взгляд на жизнь через окно генома. М.: Наука. Том 2. Очерки современной молекулярной генетики. - 2019. - 495 с.
3. Компьютеры и суперкомпьютеры в биологии Под. ред. В.Д. Лахно и М.Н. Устинина. Москва-Ижевск: Институт компьютерных исследований, 2002, 528 с.
4. Раевский О.А. Свойства химических соединений и лекарств как функции их структуры. М.: Из-во «КДУ», 2013. 376 с.
5. Дружиловский Д.С., Рудик А.В., Филимонов Д.А., Лагунин А.А., Глориозова Т.А., Пороиков В.В. Веб-ресурсы для прогнозирования биологической активности органических соединений. Известия Академии наук. Серия химическая, 2016, № 2, 384-393.
6. The Practice of Medicinal Chemistry, 4th Edition. C. Wermuth, Ed. Amsterdam a.o: Academic Press, 2015.
7. Livingstone D. Data Analysis for Chemists. Oxford et al.: Oxford University Press, 1995.
8. Han van de Waterbeemd. Structure-Property Correlations in Drug Design. Austin: Academic Press, 2003.
9. Cheminformatics Approaches to Virtual Screening. Alexandre Varnek and Alex Tropsha, Eds. Cambridge, UK: RSC Publishing, 2008.
10. Filimonov D., Dmitriev A., Rudik A., Poroikov V. Assessment of the Xenobiotics Toxicity Taking into Account Their Metabolism. In: Machine Learning and Deep Learning in Computational Toxicology, Computational Methods in Engineering & the Sciences, H. Hong (Ed.), Springer Nature. 2023, p. 21-51.
11. Tarasova O.A., Rudik A.V., Ivanov S.M., Lagunin A.A., Poroikov V.V., Filimonov D.A. Machine Learning Methods in Antiviral Drug Discovery. In: Topics in Medicinal Chemistry. Springer, Berlin, Heidelberg. 2021, 37, p. 245-279.
12. Lagunin A.A., Goel R.K., Gawande D.Y., et al. Chemo- and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: a critical review. *Natural Product Reports*, 2014, 31 (11), 1585-1611.
13. Ivanov S.M., Lagunin A.A., Poroikov V.V. (2016). In silico assessment of adverse drug reactions and associated mechanisms. *Drug Discovery Today*, 2016, 21 (1), 58-71.
14. Arnold J.W., Roach J., Azcarate-Peril M.A. Emerging Technologies for Gut Microbiome. *Trends in Microbiology*, 2016, 24 (11), 887-901.
15. Tycko J., Myer V.E., Hsu P.D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell*, 2016, 63 (3), 355-370.