



Всероссийская олимпиада студентов «Я – профессионал»

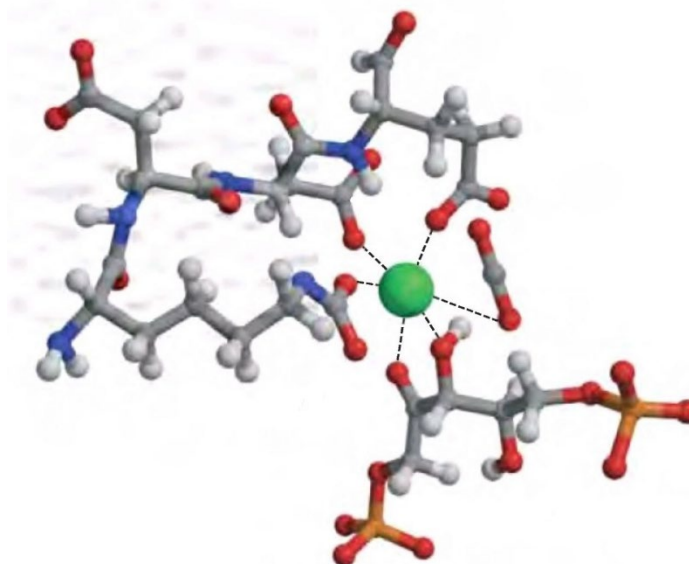
задания заключительного этапа
по направлению «Биоинженерия и биоинформатика»

Категория участия: «Бакалавриат»

Задание 1

Цикл Кальвина – многостадийный процесс, в ходе которого растительные клетки ассимилируют углекислый газ из атмосферы. Совокупность химических реакций, образующих цикл Кальвина, разделяют на три стадии: фиксация углекислого газа, восстановление и регенерация акцептора.

Первую стадию катализирует многофункциональный фермент РУБИСКО. Структура активного центра, которая непосредственно участвует в катализе, представлена на рисунке.



- 1) Расшифруйте название фермента. Какая молекула представляет собой акцептор углекислого газа? Изобразите её структуру. Какие молекулы выступают продуктами реакции? Остатки каких аминокислот и какой ион металла, непосредственно участвуют в катализе (обозначьте их на приведённом выше изображении)? Какая аминокислота не приведена на рисунке? Какие типы ферментативного катализа осуществляются при такой структуре активного центра? (3 балла)
- 2) Напишите последовательность реакций, катализируемых РУБИСКО (также необходимо указать вступающие в реакцию аминокислоты) (5 баллов).

В клетках осуществляется два пути регуляции активности РУБИСКО: с помощью РУБИСКО-активазы и при участии 2-карбоксиарабинитола-1-фосфата.

- 3) Объясните оба механизма регуляция функций РУБИСКО. Какие особенности структуры 2-карбоксиарабинитола-1-фосфата позволяют осуществлять регуляцию? (4 балла)

РУБИСКО отличается низким числом оборотов: за одну минуту при 25°C в реакцию вступают всего 3 молекулы углекислого газа. Компенсируется это значительной



представленностью фермента в хлоропластах: до 50% всех растворимых белков в хлоропластах составляет РУБИСКО.

- 4) Какой параметр в уравнении Михаэлиса-Ментен представляет собой оборотность фермента? Каков физический смысл у этого параметра (1 балл)?

Кроме углекислого газа с некоторой эффективностью фермент может фиксировать кислород, что уменьшает энергетический выход С3-растений за счет фотодыхания.

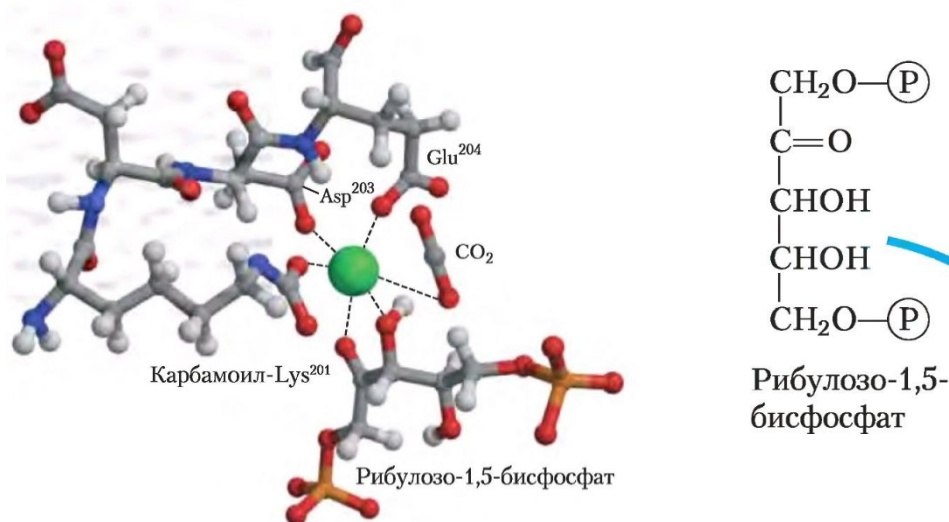
- 5) Напишите продукты реакции фотодыхания (2 балла).

Существует множество проектов, посвященных реинжинирингу фотосинтеза, направленных на повышение эффективности фотосинтеза С3-растений, например, The C4 Rice Project. Реализующие проект ученые ищут способ обойти проблему фотодыхания либо уже придуманной природой стратегией, либо альтернативными путями. Так было предложено внедрить в растения (например, рис) кластеры генов, реализующих новые биохимические пути для утилизации токсических продуктов фотодыхания, что может способствовать повышению энергоэффективности фотосинтеза С3-растений. Также способом решения данной проблемы может быть внедрение гена РУБИСКО из других организмов, поскольку бактериальная РУБИСКО обладает большей специфичностью к углекислому газу в сравнении с растительной.

- 6) Как С4-растения решили проблему фотодыхания? Чем отличается механизм фотосинтеза? у С3-растений (4 балла)?
- 7) В регенерации молекулы акцептора принимает участие фермент транскетолаза. Какой кофактор использует данный фермент? В каких других биохимических процессах задействована данная молекула (достаточно привести один пример), какой фрагмент молекулы наиболее важен для катализа? (1 балл)

Критерии оценивания задачи 1

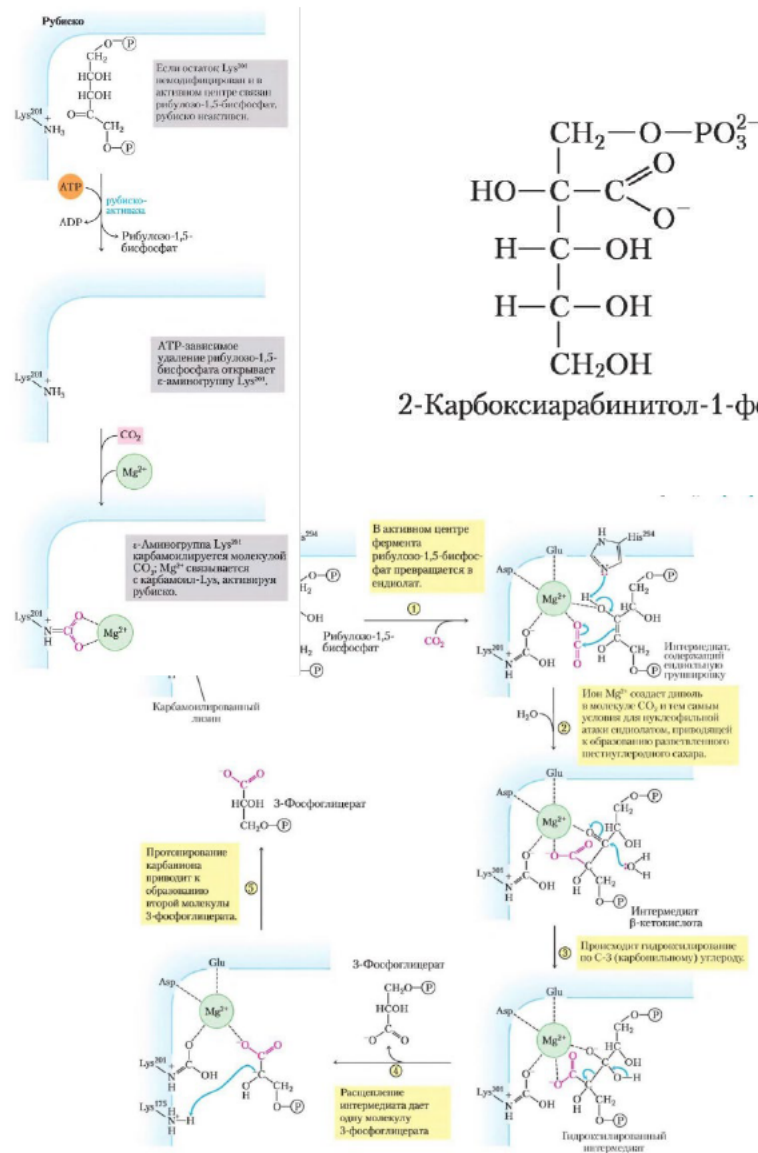
- 1) Рибулозо-1,5-бисфосфат-карбискилаза/окигеназа; рибулозо-1,5-бисфосфат; 2 молекулы фосфоглицерата; на рисунке нет гистидина 294; электрофильный и кислотно-основной катализ.





За каждый из следующих пунктов по 0,5 балла: название фермента; название акцептора и его структура, название продукта (указывать число молекул необязательно), все изображенные на рисунке аминокислоты и ион магния (если вместо карбамоиллизина написали лизин 0,25 балла при наличии остальных), отсутствующая аминокислота, механизм (за один из двух 0,25 балла). Номера аминокислот необязательны. На схеме магний обозначен зеленым шариком.

- 2) Реакция протекает через пять стадий, за каждую из которых 1 балл, за 1 ошибку в структуре интермедиатов 0,75 баллов за стадию, за две ошибки — 0,5 баллов, за 3 и более — 0 баллов. Если опущены аминокислоты, необходимые только для координации иона магния и не взаимодействующие с субстратом, решение оценивается в 1 балл за стадию.
- 3) Если рибулозо-1,5-бисфосфат попал в активный центр раньше акта карбамоилирования лизина, фермент инактивируется, РУБИСКО-активаза АТФ-зависимо удаляет из активного центра субстрат, способствуя проникновению в него углекислого газа и, как следствие, неферментативному карбамоилированию. Второй механизм регуляции направлен на карбамоилированный РУБИСКО, подавление активности которого происходит посредством конкурентного ингибирования за счет сходства структуры ингибитора со структурой переходного состояния.





За каждый из механизмов по 2 балла.

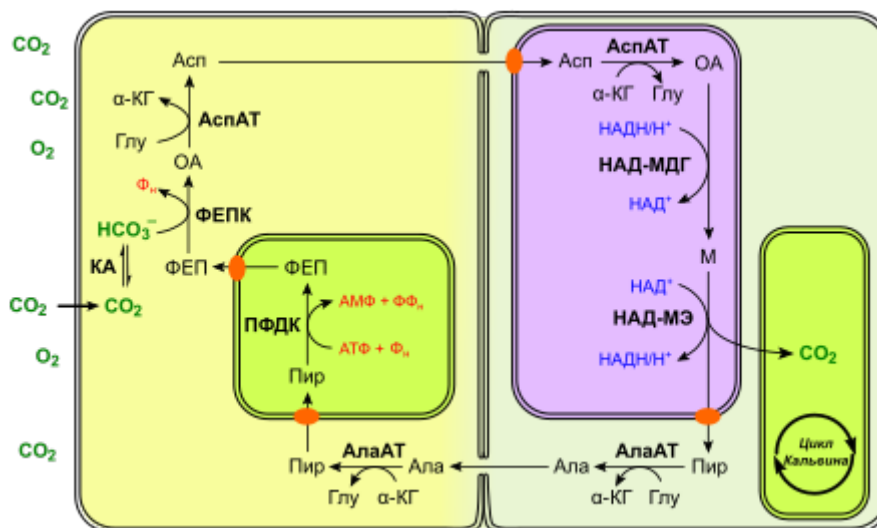
Если в первом указано именно способствование карбоамилированию 1 балл (если указан катализ или ферментативное карбоамилирование 0 баллов), если указано удаление рибулозо-1,5-бисфосфат 0,5 балла, упоминание АТФ-зависимости процесса удаления субстрата еще 0,5 балла.

Во втором 1 балл за сходство со структурой переходного состояния, 1 балл за конкурентное ингибирование.

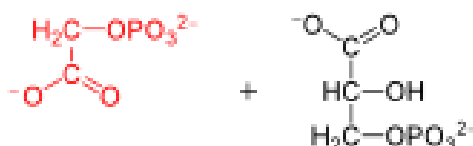
- 4) Число оборотов – это k_{cat} , физический смысл которого заключается в числе актов катализа реакции (актов превращения субстрата в продукт) в единицу времени, константа скорости первого порядка.

За определение параметра 0,25 балла, за его физический смысл 0,25 балла.

- 5) Фосфогликолат и фосфоглицерат
 За каждую молекулу по 1 баллу.
- 6) У С3-растений углекислый газ поступает в клетки мезофилла, где сразу фиксируется в цикле Кальвина, но вместе с углекислым газом в клетки поступает значительное количество кислорода, за счет чего реализуется фосфореспирация. У С4-растений цикл Кальвина протекает в клетках обкладки проводящего пучка, а углекислый газ заквываетается в клетках мезофилла, между двумя типами клеток челночно в составе аспартата или малата доставляется в клетки обкладки проводящего пучка, таким образом достигается повышение концентрации углекислого газа и значительное уменьшение концентрации кислорода, доступного для фиксации, и предупреждается фотодоыхание.



Если указано отсутствие пространственного разделения процессов в случае С3-

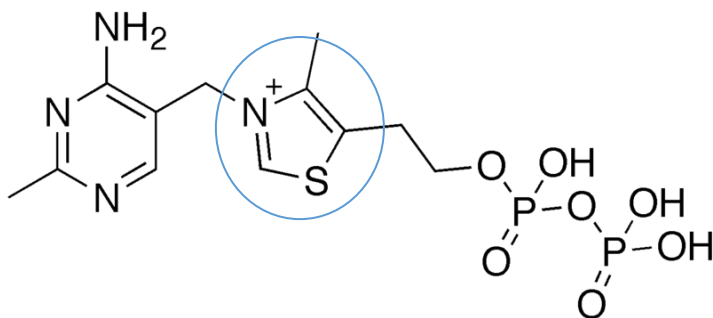


растений — 1 балл, за указание типов клеток, между которыми разделены процессы



у С4-растений — 1 балл, если указано название только одного типа клеток — 0,5 балла, если указано без названий типов — 0,25 балла. За указание малат-аспаратного челнока — 1 балл, (если указан просто челнок 0,25 балла). Если указано что углекислый газ переносится на молекулу фосфоенолпирувата с образованием оксалоацетата — 1 балл.

- 7) Тиаминпирофосфат, витамин В1, пируват-дегидрогеназный комплекс (ПДК).
За название — 0,5 балла, пример альтернативного биохимического процесса — 0,5 балла, за структуру важного для катализа фрагмента молекулы — 0,5 балла.





Задание 2

После празднования китайского нового года биоинженер Вася видит сон...

И снится ему, что он исследует две чистых линии разноцветных драконов (№1 и №2), также в его распоряжении чистая линия белых драконов (№3).

Сначала он скрещивает разноцветных драконов линии №1 с линией белых драконов №3, драконы первого поколения имеют разноцветную окраску. Во втором поколении Вася наблюдает расщепление: лишь 16% драконов белые, а 84% - разноцветные.

Тогда он скрещивает вторую чистую линию разноцветных драконов №2 с линией белых драконов №3, драконы первого поколения имеют разноцветную окраску, во втором поколении – 25 % драконов белые, а 75 % - разноцветные.

Не поверив не сходящимся результатам, Василий повторил оба эксперимента увеличив выборку в 100 000 раз. Несмотря на то, что эксперименты во сне воспроизводятся обычно плохо, расщепления воспроизвелись.

Теперь, прямо во сне, Василий ломает голову как согласовать два эксперимента. Ему и вам, дополнительно известно, что драконы – диплоидны, в основе их генетики лежит процесс мейоза. Разноцветная окраска у драконов определяется взаимодействием между доминантным аллелем Р и доминантным аллелем W, каждый из которых приводит к появлению разноцветной окраски наружных покровов.

- 1) Какими терминами можно описать подобный тип взаимодействия генов Р и W? (1 балл)
- 2) Изобразите схему первого скрещивания. Найдите генетическое расстояние между генами Р и W. Какова доля дигетерозигот среди драконов F2? (8 баллов)
- 3) Чем отличаются разноцветные чистые линии 1 и 2? Каково наиболее простое объяснение различия результатов в F2 в скрещиваниях между разными линиями? Какова доля рецессивных дигетерозигот среди драконов F2 в втором скрещивании? При обосновании приведите схему скрещивания. (6 баллов)
- 4) Как можно объяснить результаты второго скрещивания, если F1 – дигетерозигота? В чем может быть причина на уровне хромосомных перестроек? Приведите схему скрещивания. (5 баллов)

Критерии оценивания задачи 2

1. Равнозначность - 1 балл

2. Схема скрещивания

Р: PPWW x ppww
разноцветные белые
G: PW pw
F1: PW/ pw
разноцветные

2 балла



Наличие между этими генами генетического расстояния означает расположение этих генов в одной группе сцепления. Обозначим искомое расстояние за x , значит частота кроссоверных гамет будет тоже x .

F1: PW/pw

G: PW, pw (некроссоверы), Pw,pW (кроссоверы)

Тогда для F2 получаем:

	PW, $(1-x)/2$	Pw, $x/2$	pW, $x/2$	pw, $(1-x)/2$
PW, $(1-x)/2$				
Pw, $x/2$				
pW, $x/2$				
pw, $(1-x)/2$				$(1-x)^2/4$

За решетку Пенета для F2 -3 балла

Доля белых: $(1-x)^2/4 = 0,16$

$x=0,2$ или 20 сантиморганов

За верный расчет - 1 балл

Доля дигетерозигот среди всех драконов F2: Дигетерозиготы будут расположены по диагонали решетки Пенета

	PW, 0,4	Pw, 0,1	pW, 0,1	pw, 0,4
PW, 0,4				PpWw
Pw, 0,1			PpWw	
pW, 0,1		PpWw		
pw, 0,4	PpWw			

$0,4^2+0,1^2+0,1^2+0,4^2 = 0,34$

За верный расчет – 2 балла

3. Один из вариантов: линия 2 – рецессивен по одному из генов окраски, в то время как линия 1 – доминантна по обоим генам 1 балл

Таким образом, мы наблюдаем в F2 расщепление 3:1 по одному из генов

Схема скрещивания

P: ppWW x ppww

разноцветные белые

G: pW pw

F1: pW/ pw

разноцветные

F2	pW, 0,5	pw, 0,5
pW, 0,5		
pw, 0,5		pw/pw 0,25

За схему скрещивания для F1 2 балла

За решетку Пенета для F2 2 балла

Дигетрозигот в F2 во втором скрещивании-нет 1 балл

За иные разумные объяснения – не более 6 баллов



4. В этом случае

Схема скрещивания

P: PPWW x ppww
разноцветные белые
G: PW pw
F1: PW/ pw
разноцветные

у гибридов F2 не происходит кроссинговер, гены P и W наследуются сцеплено.

1 балл

F2	PW, 0,5	pw, 0,5
PW, 0,5		
pw, 0,5		pw/pw 0,25

За решетку – 2 балла

Отличия линий -в хромосомной перестройке – инверсии участка хромосомы, между генами P и W 2 балла

В этом случае в мейозе образуются хромосомы с крупными делециями и дупликациями, имеющими летальный эффект, либо хромосомы не могут разойтись из-за дупликации или делеции центромер.



Задание 3

Реализация генетической информации является одной из функций ДНК и, согласно центральной догме молекулярной биологии, состоит из двух этапов: транскрипции и трансляции. Сложность регуляции транскрипции зависит от числа вовлеченных в него белковых молекул и значительно отличается между прокариотическими и эукариотическими организмами. В эукариотических клетках функционируют несколько различных РНК-полимераз: Pol I, Pol II, Pol III и обнаруженные у растений Pol IV и Pol V. У бактерий и архей ДНК-зависимый синтез РНК выполняет единственная РНК полимераза (RNAP). Полная молекула RNAP (холофермент) состоит из корового фермента и σ -фактора, который необходим для распознавания и плавления.

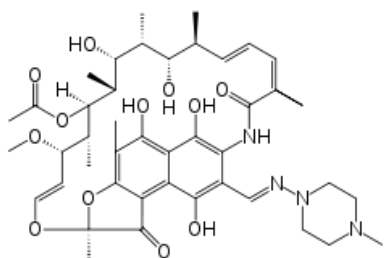
- 1) Из каких субъединиц состоит коровый фермент RNAP? Изобразите схематическую структуру холофермента RNAP (1 балл) и прокариотического промотора (1 балл), приведите консенсусные последовательности (1 балл).

Одним из способов контроля экспрессии генов у прокариот являются специализированные σ -факторы: у *E. coli* их 7, а у *Streptomyces coelicolor* – 65.

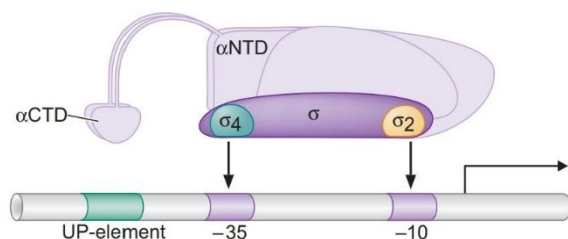
- 2) Приведите схематично структуру наиболее распространенного фактора σ^{70} *E. coli*. Изобразите с какой частью консенсусного промотора взаимодействуют структуры σ -фактора. (3 балла)
- 3) Перечислите группы генов или функции генов, ассоциированные с разными σ -факторами (5 баллов).
- 4) Что такое анти- σ -факторы и какова их функция (1 балл)?
- 5) Промоторы, узнаваемые σ^{32} -фактором, применяют в биотехнологии в качестве индуцируемых. Как его можно индуцировать? Какие гены в геноме *E. coli* контролируются такими промоторами? Каков молекулярный механизм его индукции? (4 балла).

Этап, следующий за ассоциацией холофермента RNAP и ДНК, представляет собой abortивную транскрипцию, которая может реализовываться по трем механизмам, среди которых наиболее вероятным считают «сморщивание» (scrunching), так как ДНК в рамках данного механизма сжимается за счет раздельной и достаточно незначительной транслокации обеих цепей внутри молекулы RNAP.

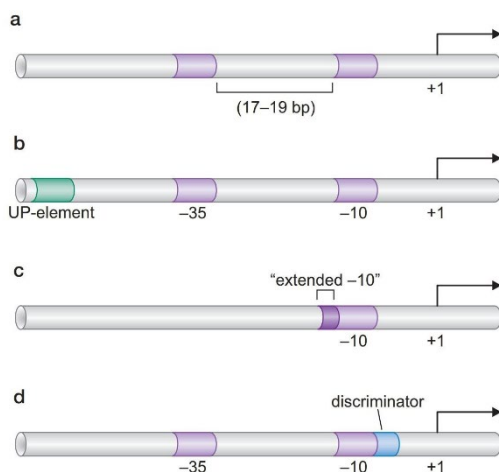
- 6) Назовите белки, которые способствуют терминации транскрипции за счет вытеснения RNAP из комплекса с ДНК. (2 балла)
- 7) Какую функцию нетранскрипционную функцию выполняет RNAP (1 балл)?
- 8) Какое вещество изображено на картинке? В чем заключается его действие на клеточные процессы? (1 балл)



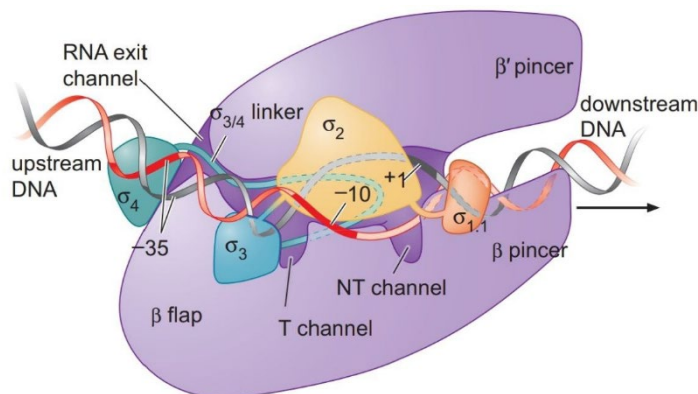
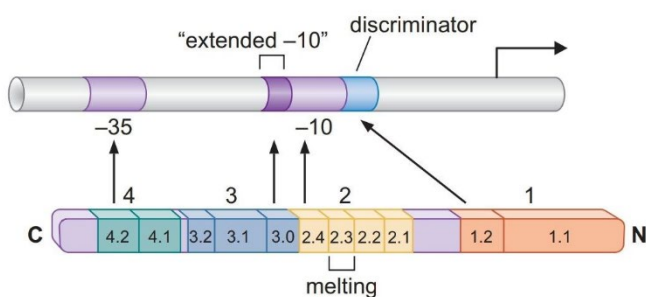
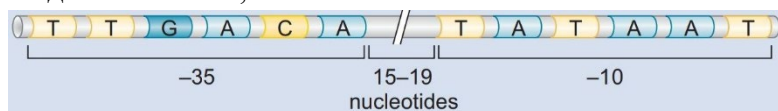
Критерии оценивания задачи 3



- 1) $\alpha 2\beta\beta'\omega$
 За субъединичный состав 0,5 балла, за структуру также 0,5 балла.
- 2) За коровый промотор 0,5 балла, за расширенный промотор 0,5 балла, за консенсусные последовательности -10 и -35 элементов 1 балл, если указана верно

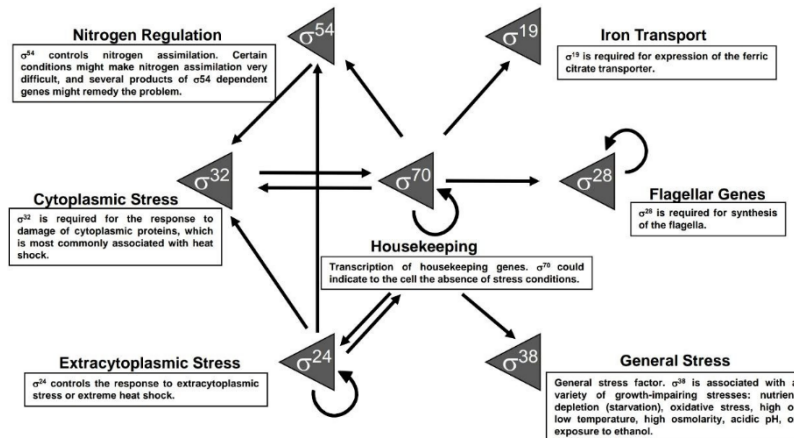


только одна из них – 0,5 балла.



- 3) σ -фактор представляет собой белок из 4 доменов, взаимодействующих с различными частями промотора и соединенных линкерами. Регионы 2 и 4 взаимодействуют с -10 и -35 элементами промотора соответственно.

За структуру сигма-фактора 1 балл, за каждый верно указанный взаимодействующий домен по 1 баллу, то есть 2 балла.



- 4) За указание σ^{70} и σ^{32} по 0,25 балла, за указание их специфичности также 0,25 балла. Для остальных факторов и их специализации по 0,5 балла.
- 5) Анти- σ -факторы представляют собой небольшие белки, которые связываются с σ -факторами и ингибируют транскрипционную активность, регулируя экспрессию генов прокариот.
 За определение 1 балла.
- 6) σ^{32} -зависимые промоторы используются в качестве термоиндуцируемых. Промотор гена любого белка теплового шока подойдет, часто используют промоторы генов шаперонов. В нормальных условиях σ^{32} связан с шапероном DnaK, что препятствует инициации транскрипции на σ^{32} -зависимых промоторах, однако при повышенных температурах (42°C) в клетках накапливается много несвернутого белка, которые связывается DnaK, приводя к высвобождению σ^{32} -фактора. В этом заключается механизм индукции.
 За выбор правильного фактора и за выбор промотора по 1 баллу. За описание механизма 2 балла (если указан анти- σ -фактор, но не раскрывается его высвобождение — 1 балл). Если будут предложены варианты с другими σ -факторами, они будут оцениваться по той же схеме, при условии, что предлагаемый вариант будет релевантным.
- 7) Rho-белок и TRCF
 За Rho-белок 1 балла, за TRCF 1 балл.
- 8) Синтез РНК-праймера для репликации некоторых фагов и плазмид, реплицирующийся по модели катящегося кольца. То есть праймазная функция.
 За эту функцию 1 балл, если будет приведена альтернативная, но релевантная функция также 1 балл.
- 9) На рисунке изображен рифампицин, антибиотик блокирующий работу РНКП за счет взаимодействия с β -субъединицей и подавления элонгации.

За название и способность блокировать синтез РНК 1 балл.



Задание 4

РНК-секвенирование позволяет провести оценку количества копий мРНК в образце. После первичной обработки данных, которая включала выравнивание на референсный геном и количественный анализ шести образцов из двух групп, получена таблица с численными значениями экспрессии генов, фрагмент которой представлен ниже.

	Группа_1_1	Группа_1_2	Группа_1_3	Группа_2_1	Группа_2_2	Группа_2_3
Ген_1	1183	1038	974	492	623	441
Ген_2	315	422	296	354	245	231
Ген_3	1245	1364	1233	1339	1208	1335
Ген_4	413	587	336	1441	1649	1596
Ген_5	190	168	198	422	485	346
Ген_6	674	581	478	1960	1782	1768
Ген_7	226	132	134	215	545	293
Ген_8	1339	1715	1072	1878	1111	1532
Размер библиотеки	10,000,000	10,500,000	9,500,000	10,200,000	9,850,000	10,400,000

В ячейках таблицы представлены значения экспрессии генов (“Ген_1 – Ген_8”) в двух группах по три технических повторности в каждой (“Группа_1” и “Группа_2”) в виде абсолютного количества прочтений, приходящихся на ген (формат - “raw counts”), а также общее количество прочтений, приходящееся на библиотеку в столбце “Размер библиотеки”. Для корректного сравнения значений экспрессии генов между библиотеками и группами, необходимо сделать нормировку значений экспрессии raw counts на размер библиотеки (CPM – counts per million) по формуле: $CPM(\text{ген}) = \frac{\text{raw counts}(\text{ген}) * 1,000,000}{\text{размер библиотеки}}$

1) Используйте данную формулу для вычисления CPM значений экспрессии для пятого гена (Ген_5) и определите, во сколько раз различается нормированная экспрессия данного гена между группами по среднему значению ($FoldChange = \frac{CPM(\text{Группа 2})}{CPM(\text{Группа 1})}$). В ответе приведите значение, округленное до второго знака после запятой (5 баллов).

2) Исходя из значений экспрессии для восьми приведенных генов в обеих группах, выявите из них самый стабильно экспрессирующийся ген (характеризующийся наименьшим среднеквадратичным отклонением), который можно будет использовать при нормировке результатов кПЦР. Вычислите для него нормированные значения экспрессии CPM в шести библиотеках; в ответе приведите среднее значение (CPM) и округлите до второго знака после запятой (2 балла). Используя вычисленные CPM значения, рассчитайте среднеквадратичное отклонение для данного гена в шести библиотеках по формуле $\sigma = \sqrt{\frac{\sum_{i=1}^n (CPM_i - \overline{CPM})^2}{N-1}}$, где N(n) – общий размер выборки; ответ округлите до второго знака после запятой (3 балла).

Метод полимеразной цепной реакции с обратной транскрипцией с детекцией результатов в режиме реального времени (ОТ-ПЦР-РВ) используют для количественного анализа экспрессии генов в клетках эукариот. Значение Cq (quantification cycle) соответствует циклу, на котором уровень флуоресценции превысил критическое значение, задаваемое настройками ПО амплификатора. Из результатов анализа калибровочной кривой получена



формула для перевода значений C_q в десятичный логарифм: $C_q = 41.5 - 3,4 * \log_{10}(copy)$, где $copy$ – количество копий мРНК в образце.

3) Для идентифицированного гена со стабильной экспрессией (Вопрос 2), известно значение его цикла выхода ОТ-ПЦР-РВ C_q (26.2) в первом образце первой группы (“Группа_1_1”). Приведите оценку цикла выхода C_q в аналогичной реакции для этого гена в первом образце первой группы, если известно, что его абсолютное количество прочтений по данным РНК-секвенирования в данной повторности составляет 400 копий (5 баллов). Исходите из допущения о том, что зависимость между результатами ОТ-ПЦР-РВ и результатами РНК-секвенирования линейная. Ответ округлите до второго знака после запятой.

4) По данным РНК-секвенирования в другой выборке образцов, для гена со стабильной экспрессией (Вопрос 2) известно среднее значение экспрессии ($expr = 430$) и среднеквадратичное отклонение экспрессии, $\sigma = 30$. Вычислите доверительный интервал цикла выхода данного гена на ОТ-ПЦР-РВ, пользуясь правилом 3σ (5 баллов). Исходите из допущения о том, что зависимость между результатами ОТ-ПЦР-РВ и результатами РНК-секвенирования линейная. В ответе приведите интервал значений 3σ , и округлите его до второго знака после запятой.

Критерии решения задачи 4

1) Для пятого гена СРМ значения экспрессии следующие: 19,00; 16,00; 20,84 (Группа 1);

41,37; 49,24; 33,27 (Группа 2); $FoldChange = \frac{СРМ(Группа 2)}{СРМ(Группа 1)} = \frac{41,29}{18,61} = 2,22$. (5 баллов).

2) Самый стабильный ген – Ген 3. При переводе в СРМ шкалу среднее значение его экспрессии в единицах СРМ получается **127,75** (2 балла). Используя вычисленные СРМ значения, рассчитываем среднеквадратичное отклонение для данного гена по

предложенной формуле $\sigma = \sqrt{\frac{10,53 + 4,66 + 4,18 + 12,45 + 26,07 + 0,38}{5}} = 3,41$ (3 балла).

3) Так как дано, что зависимость между результатами ОТ-ПЦР-РВ и результатами РНК-секвенирования линейная, значит необходимо составить пропорцию. Поскольку из решения п. 2 известно, что самый стабильный ген – Ген 3, и ему соответствует C_q (26.2), а из приведенной Таблицы количество его копий составляет 1245 (31623 копий в ОТ-ПЦР-РВ). Дано что в другой повторности количество копий – 400, следовательно получаем соотношение $400/1245 = 0,32$, значит, в другой повторности будет примерно в три раза меньше копий (10160), что соответствует C_q – **27,88** по приведенной формуле $C_q = 41.5 - 3,4 * \log_{10}(copy)$ (5 баллов).

4) Доверительный интервал по РНК-секвенированию составляет по правилу трех сигм: $[430 - 90; 430 + 90] = [340; 520]$. Вычисляем значения экспрессии ОТ-ПЦР-РВ исходя из данных п.3. – $[8636; 13208]$, что соответствует C_q [**28,12; 27,49**] (5 баллов).



Задание 5

Множественное выравнивание – это универсальный алгоритм для исследования биологических последовательностей. Множественное выравнивание можно использовать для проведения других анализов. Например, можно оценить эволюционное отношение последовательностей путем построения филогенетических деревьев или найти консервативные мотивы в последовательности при помощи различных алгоритмов. Два основных метода для поиска мотивов в последовательностях – это позиционные весовые матрицы (PWM) и скрытые марковские модели (НММ). Цель этого задания – применить эти два метода для построения профиля множественного выравнивания.

```
GTAGCTATTA
GAA-CTACTA
GTA--TATTA
GAA---ATTA
CAAGCTACTT
```

1. Для данного выравнивания постройте позиционную весовую матрицу (PWM). Выберите функцию для подсчета весов при переходе от вероятностей к весам. Чем использование матрицы с такими весами будет лучше по сравнению с использованием матрицы вероятностей? (6 баллов)
2. Постройте профиль данного выравнивания с использованием скрытых марковских моделей (НММ). В качестве узлов графа используйте состояния (совпадение, инсерция или делеция), в качестве ребер – вероятность перехода из одного состояния в другое. В структуре скрытой марковской модели отобразите только те состояния, которые имеют ненулевую вероятность. (6 баллов)
3. При помощи полученной позиционной весовой матрицы и скрытой марковской цепи посчитайте вес для данной последовательности: CTAGCTACGT. (3 балла)
4. Сравните два метода получения профиля множественного выравнивания. Какие плюсы и минусы есть у скрытых марковских моделей по сравнению с позиционной весовой матрицей? (3 балла)
5. Для каких целей могут быть использованы профили выравнивания, полученные при помощи позиционных весовых матриц или скрытых марковских моделей? Приведите не менее двух применений. (2 балла)

Критерии оценивания задачи 5

1. Правильно получена позиционная матрица частот и позиционная матрица вероятностей – 2 балла. Правильно получена позиционная весовая матрица с адекватной функцией подсчета весов – 2 балла. Правильный ответ на вопрос, чем PWM лучше PPM (более точно учитывает вероятности замен, веса аддитивны) – 2 балла.



G	4	0	0	2	0	0	0	0	0	0
A	0	3	5	0	0	0	5	0	0	4
T	0	2	0	0	0	4	0	3	5	1
C	1	0	0	0	3	0	0	2	0	0

Позиционная матрица частот

G	5	1	1	3	1	1	1	1	1	1
A	1	4	6	1	1	1	6	1	1	5
T	1	3	1	1	1	5	1	4	6	2
C	2	1	1	1	4	1	1	3	1	1

Позиционная матрица частот с псевдокаунтом

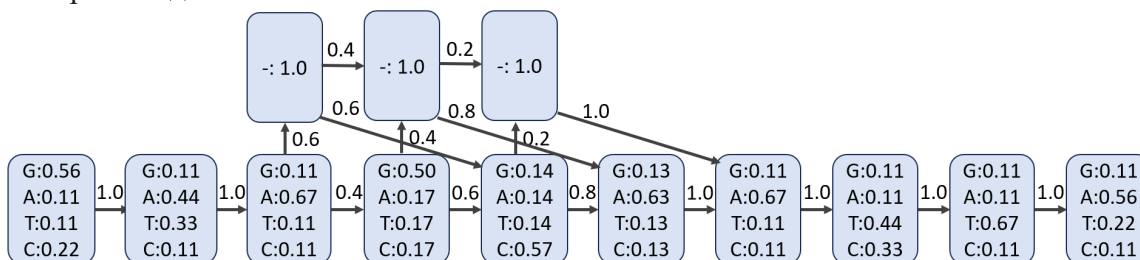
G	0.56	0.11	0.11	0.33	0.11	0.11	0.11	0.11	0.11	0.11
A	0.11	0.44	0.67	0.11	0.11	0.11	0.67	0.11	0.11	0.56
T	0.11	0.33	0.11	0.11	0.11	0.56	0.11	0.44	0.67	0.22
C	0.22	0.11	0.11	0.11	0.44	0.11	0.11	0.33	0.11	0.11

Позиционная матрица вероятностей

G	0.34	-1.29	-1.29	-0.19	-1.29	-1.29	-1.29	-1.29	-1.29	-1.29
A	-1.69	-0.31	0.11	-1.69	-1.69	-1.69	0.11	-1.69	-1.69	-0.07
T	-1.69	-0.60	-1.69	-1.69	-1.69	-0.07	-1.69	-0.31	0.11	-0.99
C	-0.60	-1.29	-1.29	-1.29	0.09	-1.29	-1.29	-0.19	-1.29	-1.29

Позиционная весовая матрица

2. Построена адекватная НММ – 6 баллов.



3. Правильно посчитан вес для последовательности, используя построенные PWM и НММ (просуммированы веса) – 4 балла
4. Правильный ответ на вопрос про плюсы и минусы НММ (+ - более точно описывает паттерн, учитывает делеции и инсерции, - - более вычислительно сложный) – 2 балла
5. Приведены адекватные примеры использования (например, поиск функциональных мотивов, поиск промоторных последовательностей) – 2 балла, по 1 баллу за пример