

Лаб. 2 часть 1

Данная часть лабораторной работы имеет более теоретический характер, нежели практический.

Формат сдачи: Jupyter Notebook с краткими ответами на вопросы от Question-a и кодом для задания на слайде 23. Или же документ с ответами и код отдельно в удобной для вас среде разработки.



Папка – здесь создать себе папку Фамилия_ИО и туда загрузить файл с ответами

Материалы:

Компо, Певзнер «Алгоритмы в биоинформатике» с. 204 (202) – 214 (212)

Перед интенсивной нагрузкой
нужна **разминка**

P.S. Гугление мать учения

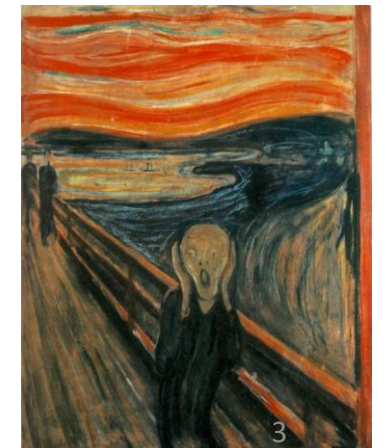
THE · SLY · FOX · AND · THE · SHY · DOG

THE · SYF · OXA · NDT · HES · HYD · OG

THE · SFO · XAN · DTH · ESH · YDO · G

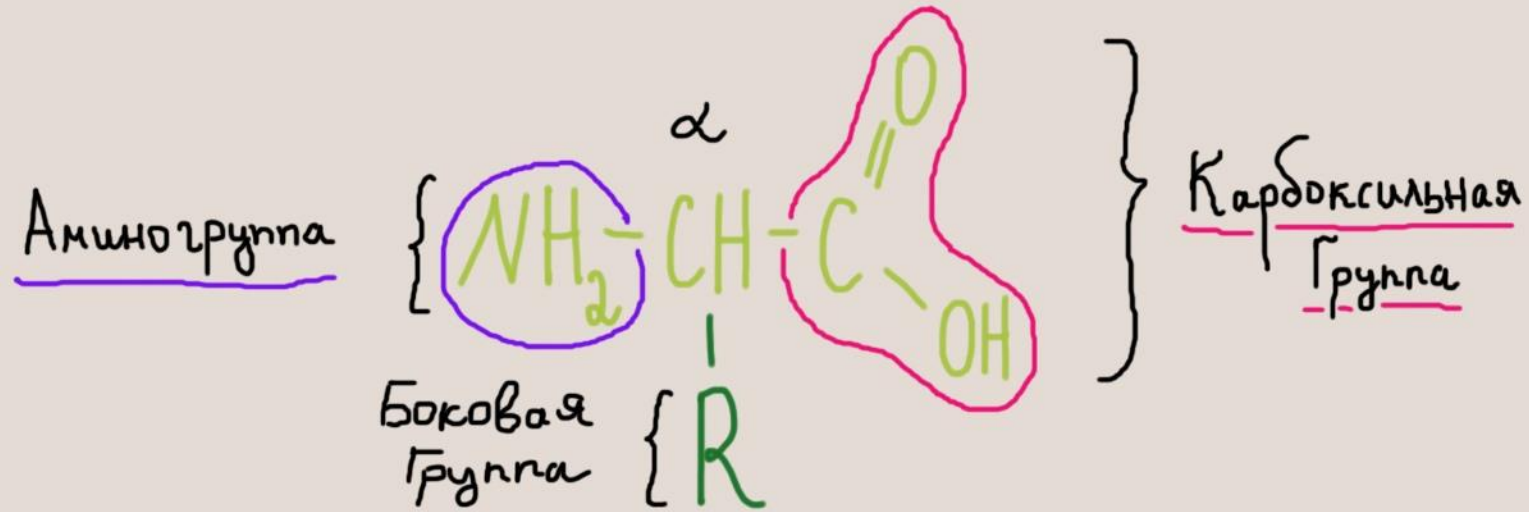
THE · FOX · AND · THE · SHY · DOG

- 1) Какое открытие представлено на данном слайде?
(назовите максимально точно)
- 2) Кто и в каком году его сделал?



Аминокислоты — это органические соединения, содержащие аминогруппу и карбоксильную группу.

Протеиногенными аминокислотами называют те, что попадают в белки в процессе трансляции. Оказывается, что в белки попадают не любые аминокислоты, а небольшой набор **α -L аминокислот**.

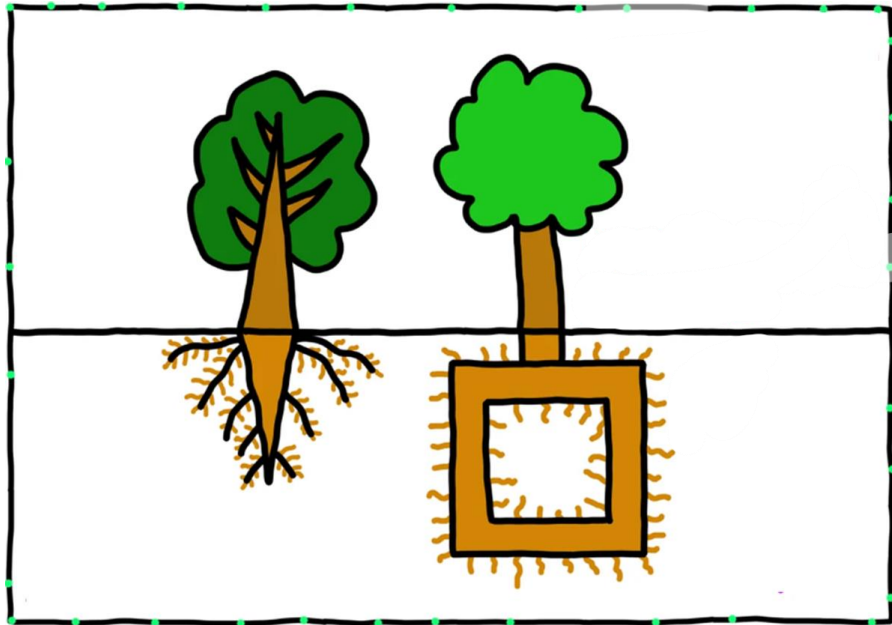


α -аминокислота

Аминогруппа у альфа-аминокислот отделена от карбоксильной группы только одним атомом углерода, так называемым альфа углеродным атомом

Альфа углеродный атом - первый углерод после карбоксильной группы

Какой частью молекулы альфа-аминокислоты отличаются друг от друга?



Протеиногенных аминокислот на текущий момент известно 22, а наиболее часто встречаются лишь 20.

Назовите те самые две (21-ю и 22-ю) аминокислоты и кодирующие их кодоны.



Секвенирование антибиотиков

Антибиотики — это лекарственные препараты, которые используют для лечения инфекций, вызванных бактериями. Они подавляют рост или уничтожают бактерии, тем самым помогая организму справиться с инфекцией. Антибиотики не действуют на вирусы, поэтому их применение неэффективно при вирусных заболеваниях, таких как простуда или грипп.

Существует несколько классов антибиотиков, которые действуют по-разному, например:

- **Пенициллины:** угнетают синтез клеточной стенки бактерий.
- **Макролиды:** нарушают синтез белка в клетках бактерий.
- **Тетрациклины:** останавливают рост бактерий, влияя на синтез белка.

Важно использовать антибиотики правильно, следуя указаниям врача, чтобы избежать развития резистентности бактерий — это состояние, когда бактерии становятся нечувствительными к антибиотикам.

from Chat GPT

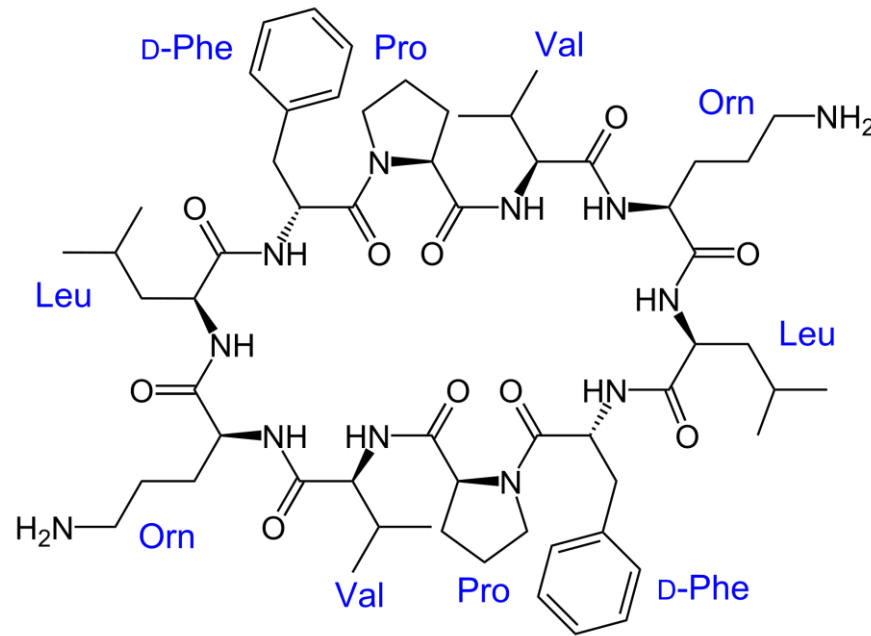
Пептиды — это короткие цепочки аминокислот, соединённые между собой пептидными связями. Аминокислоты — это строительные блоки белков, и когда несколько аминокислот соединяются в цепочку, образуются пептиды. В зависимости от длины цепи их различают на:

- **Олигопептиды:** короткие цепочки из 2–20 аминокислот.
- **Полипептиды:** более длинные цепочки, состоящие из 20–50 аминокислот.

Если число аминокислот в цепочке превышает 50, такие молекулы чаще называют **белками**, хотя строгой границы между полипептидами и белками нет.

Из-за своих небольших размеров пептиды более подвижны, чем белки, и могут легко проникать через клеточные мембраны, что делает их перспективными для использования в медицине и косметологии.

Идея: Часть антибиотиков является пептидами.



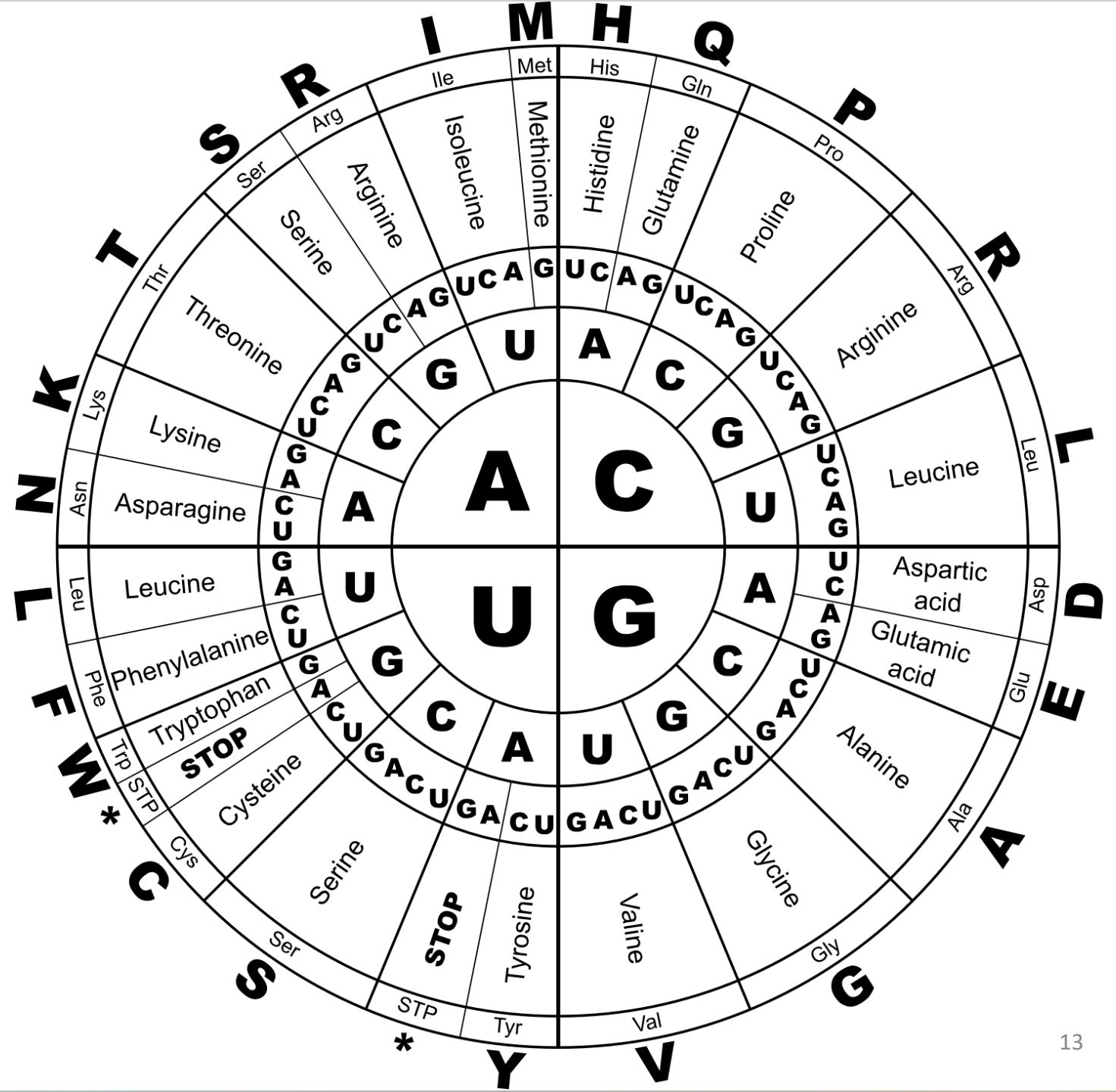
Грамицидин (*gramicidin*, *gramicidinum*, Грамицидин С) — антибиотик тиротрициновой группы.

Продуцируется споровой палочкой *Bacillus brevis* var. *G.-B.*^[1].

Один из двух первых советских антибиотиков, созданных в 1942 году (второй — **крустозин**)^[2].

Центральная догма молекулярной биологии гласит, что «ДНК производит РНК, РНК делает белок». Согласно центральной догме, ген из генома сначала **транскрибируется** в цепь РНК, состоящую из четырех **рибонуклеотидов**: аденина, гуанина, цитозина и урацила. Нить РНК может быть представлена как цепь РНК, состоящая из четырехбуквенного алфавита {A, C, G, U}. Вы можете думать о геноме как о большой поваренной книге, и в этом случае ген и его РНК-транскрипт образуют рецепт в этой поваренной книге. Затем транскрипт РНК **транслируется** в аминокислотную последовательность белка.

Genetic code wheel



Последовательность аминокислот грамицидина С (most likely)

Val-Gly-Ala-Leu-Ala-Val-Val-Val-Trp-Trp-Trp-Leu



Сколько различных цепей ДНК длиной 36 транскрибируется и транслируется в грамицидин С (не считая стоп кодон)?

Прим.

«потенциально» транскрибируется и транслируется

Есть три разных способа разделить цепь ДНК на кодоны для трансляции по одному, начиная с трех первых начальных позиций цепи. Эти разные способы называются **рамками считывания**.

2. **Рамка считывания:** Рамка считывания начинается с первого кодона, который соответствует аминокислоте (обычно это кодон AUG, который кодирует метионин). Рамка считывания определяет, как последовательность нуклеотидов (триплеты) будет интерпретироваться рибосомой. Считывание происходит с 5'-конца к 3'-концу мРНК.

3. **Кодоны:** Каждый кодон состоит из трех нуклеотидов. Например, в мРНК последовательность "AUG GUC CUA" будет читаться как:

- AUG - метионин (начало)
- GUC - валин
- CUA - лейцин

Примеры рамок считывания

При наличии одной и той же мРНК можно выделить три разные рамки считывания, если сместить начальную точку считывания на один или два нуклеотида. Это может привести к образованию совершенно разных белков.

Например, для последовательности мРНК "AUG UUC CCG" (считая от 5' до 3'):

- 1-я рамка: AUG UUC CCG (метионин - фенилаланин - пролин)
- 2-я рамка: UGU UCC CG (цистеин - серин - стоп)
- 3-я рамка: GUU CCG (валин - пролин)

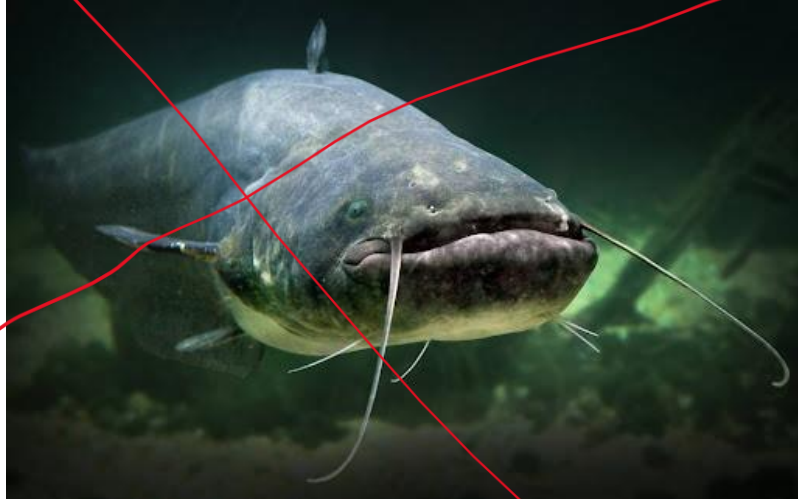
Грамицидин С – циклический пептид



Тогда сколько различных аминокислотных последовательностей может его кодировать?

Нерибосомные пептиды (NRPs)

Обычно трансляция белков происходит на рибосоме и если ее заблокировать, то производство большинства белков прекратится. Но не некоторых пептидов. Такие пептиды называются нерибосомными и синтезируются NRP-синтетазой.



NRP-синтетаза

Проблемы

- 1) Мы не можем вывести последовательность нерибосомных пептидов из ДНК, т.к. они синтезируются другим способом
- 2) Многие пептиды еще и зациклены

Решение

Использовать масс-спектрометр

Молекулярная масса

Дальтон (сокращенно **Да**) – это единица измерения атомной массы в молекулярном масштабе. Один дальтон эквивалентен одной двенадцатой части массы углерода-12 и имеет значение примерно $1,66 \cdot 10^{-27}$ кг. **Моноизотопная масса** молекулы равна сумме масс атомов в этой молекуле с использованием массы наиболее распространенного **изотопа** для каждого элемента (см. таблицу ниже).

Примечание

- Все изотопы одного элемента имеют одинаковый заряд ядра, различаясь лишь числом нейтронов.
- Массовое число A – суммарное количество нейтронов и протонов в ядре.

Задача секвенирования циклопептидов

Пока для простоты будем считать, что масс-спектрометр разрывает копии циклического пептида по любым возможным связям, так что результирующий экспериментальный спектр содержит массы всех возможных линейных фрагментов пептида, называемых **субпептидами**. Например, циклический пептид NQEL имеет 12 возможных субпептидов: N, Q, E, L, NQ, QE, EL, LN, NQE, QEL, ELN и LNQ. Мы также предполагаем, что субпептиды могут встречаться более одного раза, если аминокислота встречается в пептиде несколько раз (например, ELEL также имеет 12 субпептидов: E, L, E, L, EL, LE, EL, LE, ELE, LEL, ELE и LEL).



Упражнение. Сколько субпептидов имеет циклический пептид длины n ?

Теоретический спектр циклического пептида *Peptide*, обозначаемого *Cyclospectrum(Peptide)*, представляет собой совокупность всех масс его субпептидов в дополнение к массе 0 и массе всего пептида, причем массы упорядочены от наименьшей к наибольшей. Будем считать, что теоретический спектр может содержать повторяющиеся элементы, как в случае NQEL (показан ниже), где **NQ** и **EL** имеют одинаковую массу.

	L	N	Q	E	LN	NQ	EL	QE	LNQ	ELN	QEL	NQE	NQEL
0	113	114	128	129	227	242	242	257	355	356	370	371	484

Таблица молекулярных масс аминокислот

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186



Постройте теоретический спектр для
Грамицидина С

Аминокислотная последовательность: Val-Gly-Ala-Leu-Ala-Val-Val-Val-Trp-Trp-Trp-Leu