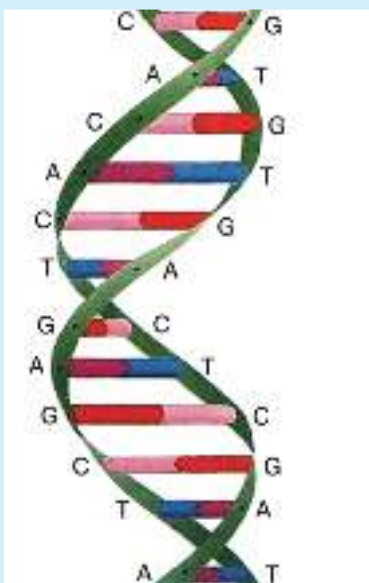


Биоинформатика.

Как изучают белки



На этой модели видно, что аденин (А) всегда связан с тиминном (Т), а цитозин (С) — с гуанином (G). Две цепочки молекулы ДНК закручены в спираль — наподобие винтовой лестницы. Эту модель (она называется «двойная спираль») придумали в 50-х гг. XX в. Дж. Уотсон и Ф. Крик

Ты уже знаешь, генетический код в основном устроен так же, как и код в наших задачах: есть текст в 4-буквенном алфавите (А, С, G, Т); тройки букв (тройки нуклеотидов молекулы ДНК) кодируют остатки — звенья молекул белков.

Участки ДНК, которые кодируют белки (они называются *генами*), ограничены специальными тройками. Первая тройка в любом гене — АТG (она называется *старт-кодоном*). Кодон АТG — это кодон остатка метионина (такой остаток обозначается буквой М), поэтому этот кодон встречается и внутри генов. Последняя тройка в каждом гене — это ТAA, TAG или TGA. Эти кодоны называются *стоп-кодонами*, они играют такую же роль, как точки в предложении. Стоп-кодоны не кодируют никакого остатка, внутри генов они не встречаются.

Ты уже знаком с *таблицей генетического кода*. В ней для каждого остатка приведены его кодоны, при этом один остаток, как правило, может кодироваться несколькими разными кодонами. На

форзаце в конце учебника приведена и обратная таблица: для каждого кодона указан соответствующий остаток.

Мы описали основной способ кодирования белков. Бывают и более сложные случаи: например, гены могут перекрываться, т. е. получается как бы одна шифровка, наложенная на другую. Такое явление встречается в ДНК вирусов. Похожие ситуации вам встретятся в задачах на шифрование.

Современные технологии позволяют определять последовательность нуклеотидов в молекулах ДНК. Сегодня уже полностью расшифрованы ДНК человека, шимпанзе, мыши, сотен других животных и растений и десятков тысяч бактерий. Теперь учёным важно понять, для чего служат различные фрагменты ДНК, и прежде всего выделить участки, кодирующие белки.

Иногда учёным приходится иметь дело с данными, в которых есть ошибки, — в ходе эксперимента некоторые буквы могут потеряться, а могут и ошибочно появиться новые буквы. Подобные задачи на шифрование тебе уже встречались. При их решении ты использовал то, что не все тройки допустимы. В реальном же кодировании белков это не так — каждая тройка имеет значение. Поэтому, чтобы выделять кодирующие участки, учёным-биологам приходится использовать более сложные методы.

218

Иногда в одной шифровке разведчику удаётся передать два сообщения, наложив «одно поверх другого». Для этого он так составляет шифровку, чтобы сначала читалось первое сообщение, а с некоторого места (например, со второй буквы шифровки) читалось второе сообщение. Вот пример такой шифровки, в которой содержатся с перекрытием 2 слова из 4 букв. Расшифруй и запиши в тетрадь эти слова, рядом с каждым словом запиши его шифровку.

С А А С G A A G T T A A A C A T

219

Пользуясь шифровальной таблицей, напиши 6 разных шифровок слова МОЩЬ. Сколько всего существует таких шифровок? Объясни свой ответ.

220

Пользуясь своей шифровальной таблицей, определи, сколькими способами можно зашифровать слово ЖАДНЫЙ. Запиши все различные шифровки этого слова.

221

На форзаце в конце учебника приведена обратная таблица генетического кода — для каждого кодона указан соответствующий остаток. Обрати внимание, что главную роль в определении остатка, соответствующего кодону, играют первые 2 буквы кодона. В одних случаях третья буква может быть любой (так, например, для *пролина* годится любой кодон, первые 2 буквы которого — СС). В других — третьи буквы А или Г соответствуют одному остатку, а С или Т — другому. Например, ААА и ААГ кодируют *лизин*, а ААС и ААТ — *аспарагин*. Из этих правил есть только два исключения. Найди их в таблице генетического кода.

222

В одной из шифровок слова РЫСЬ вычеркнули 3 буквы (не обязательно идущие подряд), и получилась одна из шифровок слова ПЕЛ. Найди и запиши такую шифровку слова РЫСЬ, подчеркни в ней те буквы, которые вычеркнули, чтобы получить шифровку слова ПЕЛ.

223

Пользуясь своей шифровальной таблицей, придумай и запиши слово, которое можно зашифровать более чем 10 способами. Запиши любые 5 из возможных шифровок этого слова.

224

Запиши предложение, содержащееся в следующей шифровке (слова, как обычно, разделены тройкой латинских букв, кодирующей пробел):

T A G T A T A C T G T C C T A T T T -
 C C A G C G C G G A G G T A C C T A -
 T G A G G T A G T G C T G T A A C T T -
 G C T A A C C C T C T T C G C T A A -
 G C G A T C T A G A A A A G G G C -
 C C T A T G A C

225

Сколько разных чисел можно получить, переставляя цифры числа **9854**?

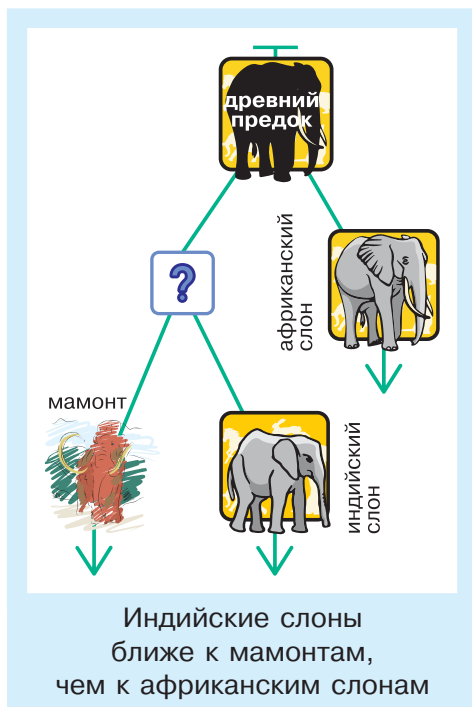
Биоинформатика. Сравнение белков

У живых организмов различаются наборы белков. Наборы белков у двух людей (у двух слонов, двух бактерий) очень похожи, только мелкие различия определяют, например, различную форму уха у разных людей или разный цвет хвоста у двух кроликов.

Наборы белков у организмов разных видов разные, но чем более родственны эти виды, тем более похожи наборы белков. Например, сходство между белками человека и шимпанзе достигает 99 %.

Верно и обратное: если два вида организмов имеют похожие наборы белков, то эти виды являются родственными, т. е. происходят от какого-то одного вида древних животных. Пользуясь методами сравнения белков, учёные восстанавливают родственные связи и происхождение различных видов животных и растений. Например, недавно стало известно, что, судя по некоторым белкам, индийские слоны более близки к вымершим мамонтам, чем к современным африканским слонам.

Как учёные проводят сравнение белков? Как вы теперь знаете, белки — это последовательности аминокислотных остатков, их можно представить в виде длинных слов — последовательностей букв в 20-буквенном алфавите. Поэтому задачу можно сформулировать так: какие два слова считаются близкими, похожими, а какие — далёкими?



Превращение слов

Вы наверняка играли в игру, в которой одно слово надо превратить в другое, заменяя на каждом шагу только одну букву. Примерно такой игрой мы сейчас займёмся.

В отличие от тех игр, которыми мы занимались раньше, в эту игру можно играть одному. Цель игры — за наименьшее число шагов превратить одно слово в другое, соблюдая такое правило:

Правило превращения слов

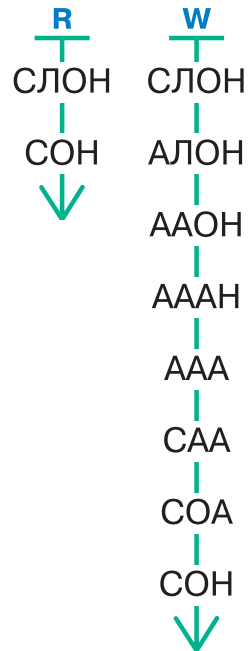
На каждом ходу можно выполнить одно из трёх действий:

- заменить одну букву на другую;
- вставить одну букву;
- удалить одну букву.

Конечно, по этим правилам одно слово можно превратить в другое разными способами. Например, из слова СЛОН можно сделать слово СОН за один ход, а можно и за 7 ходов. Последовательности таких превращений помещены снизу. Но цель нашей игры — найти самое короткое превращение, поэтому мы выбираем превращение R.

Давайте усложним игру — присвоим каждому ходу цену.

Ход	Цена
Замена гласной на гласную	1
Замена согласной на согласную	2
Замена гласной на согласную или наоборот	3
Замена Ъ или Ь на другую букву или наоборот	4
Вставка или удаление одной буквы	5



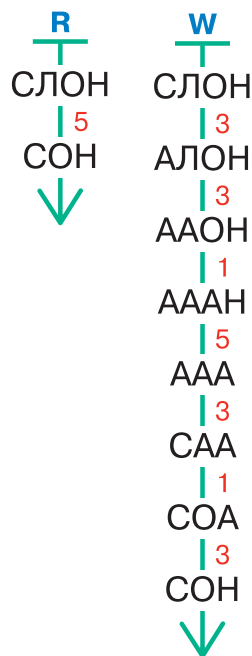
Вычислим стоимость каждого превращения: стоимость R — 5 баллов, стоимость W — 19 баллов. Чтобы было проще подсчитывать баллы, можно записывать цену ходов прямо на схеме превращения, как показано справа (примерно так мы записывали вес рёбер на деревьях и графах).

Теперь можно сформулировать новую цель игры: построить для данной пары слов самое дешёвое превращение одного слова в другое. Стоимость такого превращения можно назвать *расстоянием между словами*.

Учёные сравнивают слова-белки примерно так же: ищут для них «самую дешёвую» цепочку преобразований (при подходящих стоимостях замен и удалений). Например, замена остатка на остаток с близкими химическими свойствами стоит дешевле, чем его замена на непохожий остаток (как в нашей таблице дешевле заменить гласную на гласную, чем гласную на согласную). Удаление и вставка остатков стоят дороже, чем замены. Конечно, полная таблица «удалённости» разных остатков устроена гораздо сложнее, чем наша.

Результаты сравнения белков биоинформатики обычно изображают в виде специальной схемы — *выравнивания*. Вот, например, выравнивание слов ПАПКА и ПАПАХА и соответствующая цепочка превращения:

Сейчас учёным известны сотни тысяч последовательностей белков. Получив новый белок, учёные прежде всего стараются найти известные белки, похожие на него. Не стоит забывать, что биоинформатика имеет дело с очень длинными словами, работать с которыми можно только с помощью компьютеров. Для сравнения белков используются компьютерные программы, которые для любых двух белков могут определить их наилучшее (т. е. имеющее наименьшую стоимость) выравнивание. Такие программы



и базы данных, содержащие последовательности известных белков, доступны всем через Интернет.

Выравнивание позволяет предположить, какая последовательность остатков была у общего предка двух видов живых существ. Вот, например, выравнивание двух фрагментов белков: один из них — белок человека, а другой — белок мартишки:

```
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLS
VLSPADKSNVKAAWGKVGSHAGDYGAEALERMFLS
```

Значит, можно предположить, что наш общий предок имел такую белковую последовательность:

```
VLSPADK + NVKAAWGKVG + HAG + YGAEALERMFLS
```

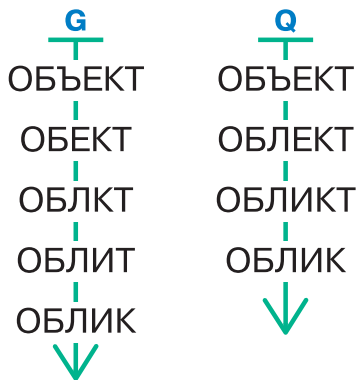
(плюсами помечены места, буквы в которых нам неизвестны: T или S, A или S, E или D).

226

Построй какое-нибудь (необязательно самое дешёвое) превращение слова ТАНКИСТ в слово ТУРИСТ. Напиши последовательность своего превращения и вычисли его стоимость.

227

Вот последовательности G и Q превращений слова ОБЪЕКТ в слово ОБЛИК. Вычисли стоимость каждого превращения и укажи более дешёвое.



228

Построй какое-нибудь превращение (необязательно самое дешёвое) слова ВРАЗБРОС в слово ВРАЗРЕЗ. Напиши последовательность этого превращения и вычисли его стоимость.

229

Построй превращение слова КОНТАКТ в слово КОМПЛЕКТ, которое стоит 10 баллов.

КОНТА – КТ
КОМПЛЕКТ

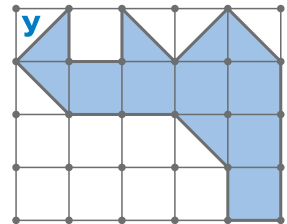


Чтобы построить превращение с заданной стоимостью, удобно сначала *выровнять* заданные слова: расположить одно слово под другим так, чтобы одинаковые буквы (конечно, если они есть) стояли друг под другом. При необходимости можно раздвигать буквы, вставляя чёрточку.

Как именно нужно расположить остальные (неодинаковые) пары букв, зависит от того, какова заданная стоимость. Например, указанное выравнивание предполагает замену Н на М (2 балла), замену Т на П (2 балла), замену А на Л (3 балла) и вставку Е (5 баллов) — всего 12 баллов. Это не та стоимость, которая требуется в задаче. Значит, надо попытаться поставить чёрточку в другом месте и снова подсчитать баллы.

230

Нарисуй, как разрезать многоугольник У, чтобы получились два одинаковых многоугольника на сетке.

**231**

За какое наименьшее число вопросов можно наверняка отгадать натуральное число, меньше 1000, если на вопросы можно отвечать только «да» и «нет»?

232

Построй превращение слова СЪЕДОБНЫЙ в слово ОГРОМНЫЙ, которое стоит 15 баллов.

233

Построй превращение слова ПОЛЬСКИЙ в слово КОНСКИЙ, которое стоит меньше 10 баллов.

234

Построй превращение слова ПОДОСИНОВИК в слово ПОДБЕРЁЗОВИК, которое стоит 11 баллов.

Подумай, существует ли более дешёвое превращение слова ПОДОСИНОВИК в слово ПОДБЕРЁЗОВИК. Объясни свой ответ.