



Математика для школьников 10 – 11 классов (отборочный этап) Решение задачи 9. Расширенный алфавит

1. Избыточность природного кода $4^3/20 = 64/20 = 3,2$ (число вариантов трехбуквенного слова, записанного алфавитом из четырех букв, отнесенное к числу шифруемых аминокислот).

Для того, чтобы зашифровать 20 аминокислот расширенным кодом, хватит всего двух букв, поскольку число вариантов двухбуквенного слова, записанного алфавитом из шести букв, больше, чем число кодируемых аминокислот: $6^2 = 36 > 20$. При этом избыточность составляет $36/20 = 1,8$, что в $3,2/1,8 \approx 1,8$ раз меньше, чем для природного кода.

2. При стандартном способе кодирования используются 4 нуклеотидных буквы, то есть, для произвольной последовательности длиной n букв возможно 4^n вариантов текста, что отвечает объему информации, равному $I_4 = \log_2 4^n = 2n$ бит информации.

В то же время, при записи последовательности такой же длины с использованием расширенного алфавита из шести нуклеотидов мы получаем один из 6^n вариантов. В этом случае объем информации равен $I_6 = \log_2 6^n = \log_2 2^n 3^n = n + n \log_2 3 \approx 2,58495n \approx 2,58n$.

Тогда увеличение объема записанной информации будет в $I_6/I_4 = 2,58n/2n = 1,29$ раз.