

Задача по расшифровке генетического кода

М.С.Гельфанд

Перед Вами данные, доступные исследователям генетического кода на начало 1960 г., непосредственно перед опытом по прямому анализу аминокислот, кодируемых кодонами (нуклеотидными триплетами). Эти данные слегка идеализированы: удалены недостоверные данные и добавлено небольшое количество данных, полученных при проверки уже открытого кода. Тем не менее ситуация сильно напоминает ту, в которой оказались первые исследователи, пытавшиеся расшифровать генетический код.

Представлены четыре типа данных. Первый – последовательности полипептидов, кодируемые регулярной нуклеотидной последовательностью с известной структурой. Второй – смеси полипептидов, кодируемые нерегулярной нуклеотидной последовательностью с известным соотношением нуклеотидов (обратите внимание: эти данные неполные, для нуклеотидной последовательности могут быть представлены не все пептиды, получающиеся из нее.) Третий тип данных – мутации в генах, вызываемые азотистой кислотой. Такие мутации могут менять А на G и С на U, приводя к изменениям в кодонах и, соответственно, в аминокислотных последовательностях, кодируемых данными генами. Четвертый вариант – результаты спонтанных (случайных) мутаций (эти данные могут быть использованы для проверки правильности расшифровки генетического кода.)

Ваша задача – реконструировать таблицу генетического кода, выяснив соответствие между кодонами и аминокислотами. Представленные данные не позволяют расшифровать генкод полностью, но Вы должны стремиться разгадать значение как можно большего числа кодонов.

Данные:

1а. Следующие регулярные полинуклеотиды дают следующие полипептиды

	короткое название	нуклеотидная последовательность	аминокислотная последовательность(и)
1	polyU	...UUUUUUUUUUUUU...	...-Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe-...
2	polyA	...AAAAAAAAAAAAA...	...-Lys-Lys-Lys-Lys-Lys-Lys-Lys-Lys-Lys-...
3	polyC	...CCCCCCCCCCCCC...	...-Pro-Pro-Pro-Pro-Pro-Pro-Pro-Pro-Pro-...
4	polyUC	...UCUCUCUCUCUCU...	...-Leu-Ser-Leu-Ser-Leu-Ser-Leu-Ser-Leu-...
5	polyUG	...UGUGUGUGUGUGU...	...-Val-Cys-Val-Cys-Val-Cys-Val-Cys-Val-...
6	polyAC	...ACACACACACACA...	...-Thr-His-Thr-His-Thr-His-Thr-His-Thr-...
7	polyAG	...AGAGAGAGAGAGA...	...-Arg-Glu-Arg-Glu-Arg-Glu-Arg-Glu-Arg-...
8	polyUUAC	...UUACUUACUUACU...	...-Leu-Leu-Thr-Tyr-Leu-Leu-Thr-Tyr-Leu-...
9	polyUAUC	...UAUCUAUCUAUCU...	...-Tyr-Leu-Ser-Ile-Tyr-Leu-Ser-Ile-Tyr-...

1b. Следующие регулярные полинуклеотиды кодируют следующие регулярные полипептиды

10	polyAAG	...AAGAAGAAGAAGA...	...-Arg-Arg-Arg-Arg-Arg-Arg-Arg-Arg-Arg-... ...-Lys-Lys-Lys-Lys-Lys-Lys-Lys-Lys-Lys-... ...-Glu-Glu-Glu-Glu-Glu-Glu-Glu-Glu-Glu-...
11	polyUAC	...UACUACUACUACU...	...-Leu-Leu-Leu-Leu-Leu-Leu-Leu-Leu-Leu-... ...-Thr-Thr-Thr-Thr-Thr-Thr-Thr-Thr-Thr-... ...-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr-...
12	polyGUA	...GUAGUAGUAGUAG...	...-Val-Val-Val-Val-Val-Val-Val-Val-Val-... ...-Ser-Ser-Ser-Ser-Ser-Ser-Ser-Ser-Ser-...
13	polyAUC	...AUCAUCAUCAUCA...	...-Ser-Ser-Ser-Ser-Ser-Ser-Ser-Ser-Ser-... ...-Ile-Ile-Ile-Ile-Ile-Ile-Ile-Ile-Ile-... ...-His-His-His-His-His-His-His-His-His-...
14	polyGAU	...GAUGAUGAUGAUG...	...-Asp-Asp-Asp-Asp-Asp-Asp-Asp-Asp-Asp-... ...-Met-Met-Met-Met-Met-Met-Met-Met-Met-...
15	polyUUG	...UUGUUGUUGUUGU...	...-Leu-Leu-Leu-Leu-Leu-Leu-Leu-Leu-Leu-... ...-Val-Val-Val-Val-Val-Val-Val-Val-Val-... ...-Cys-Cys-Cys-Cys-Cys-Cys-Cys-Cys-Cys-...
16	polyCAA	...CAACAACAACAAC...	...-Thr-Thr-Thr-Thr-Thr-Thr-Thr-Thr-Thr-... ...-Asn-Asn-Asn-Asn-Asn-Asn-Asn-Asn-Asn-... ...-Gln-Gln-Gln-Gln-Gln-Gln-Gln-Gln-Gln-...
17	polyUUC	...UUCUUCUUCUUCU...	...-Ser-Ser-Ser-Ser-Ser-Ser-Ser-Ser-Ser-... ...-Leu-Leu-Leu-Leu-Leu-Leu-Leu-Leu-Leu-... ...-Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe-Phe-...

1с. Два регулярных полинуклеотида, которые не дают длинных пептидов (в основном продуцируются пептиды длиной в три аминокислоты):

18	polyGUA	...GUAAGUAAGUAAG...
19	polyGAUA	...GAUAGAUAGAUAG...

2. Следующие нерегулярные полинуклеотиды (фракция преобладающего нуклеотида 80%) продуцируют нерегулярные полипептиды указанного состава:

	преобладающий нуклеотид.	остальные нуклеотиды.	преобладающая аминокислота	редкая(ие) аминокислота(ы)	очень редкая(ие) аминокислота(ы)
1	U	C	Phe	Ser, Leu	Pro
2	U	A	Phe	Leu, Ile, Tyr	Asn
3	U	G	Phe	Cys, Val, Leu	Gly, Trp
4	A	U	Lys	Asn, Ile	Leu, Tyr
5	A	G	Lys	Arg, Glu	Gly
6	A	C	Lys	Asp, Gln, Thr	His, Pro

3. Мутации в белке оболочки вируса табачной мозаики, вызванные превращениями A→G и C→U:

Было	Стало
Ala	Val
Asp	Gly
Glu	Gly
Ile	Val, Met
Lys	Arg
Met	Val
Asn	Ser
Pro	Leu, Ser
Gln	Arg
Arg	Gly
Ser	Gly, Leu, Phe
Thr	Ala, Met, Ile
Tyr	Cys

Та же таблица в другом формате:

Было	Стало
Thr	Ala
Tyr	Cys
Ser	Phe
Glu, Arg, Asp, Ser	Gly
Thr	Ile
Pro, Ser	Leu
Thr, Ile	Met
Lys, Gln	Arg
Asn, Pro	Ser
Ile, Met, Ala	Val

4. Сводные данные по спонтанным мутациям в различных белках (триптофан-синтаза *Escherichia coli* и гемоглобин человека)

Было	Стало
Ala	Asp, Val, Glu
Cys	Gly
Asp	Gly, Ala, Asn
Glu	Gln, Val, Stop, Gly, Ala, Asp, Lys
Phe	Leu
Gly	Val, Glu, Arg, Asp, Cys
His	Tyr, Arg, Asp, Asn
Ile	Thr, Ser, Asn
Lys	Glu, Asn, Gln
Leu	Arg, Phe
Asn	Lys, Ser
Pro	Gln
Gln	Glu, Arg
Arg	Ile, Gly, Thr, Ser
Ser	Arg, Leu, Phe, Thr
Thr	Ile, Lys, Asn, Ser
Val	Ala, Gly, Asp, Glu
Tyr	Cys

Существуют два общепринятых формата таблицы генетического кода . Я уже заполнил наиболее очевидные клетки.

аминоксилота	трехбуквенное обозначение	однобуквенное обозначение	кодон(ы)
Аланин	Ala	A	
Аргинин	Arg	R	
Аспарагин	Asn	N	
Аспартат (аспарагиновая кислота)	Asp	D	
Валин	Val	V	
Гистидин	His	H	
Глицин	Gly	G	
Глутамат (глутаминовая кислота)	Glu	E	
Глутамин	Gln	Q	
Изолейцин	Ile	I	
Лейцин	Leu	L	
Лизин	Lys	K	AAA
Метионин	Met	M	
Пролин	Pro	P	CCC
Серин	Ser	S	
Тирозин	Tyr	Y	
Треонин	Thr	T	
Триптофан	Trp	W	
Фенилаланин	Phe	F	UUU
Цистеин	Cys	C	

UUU Phe	UCU	UAU	UGU
UUC	UCC	UAC	UGC
UUA	UCA	UAA	UGA
UUG	UCG	UAG	UGG
CUU	CCU	CAU	CGU
CUC	CCC Pro	CAC	CGC
CUA	CCA	CAA	CGA
CUG	CCG	CAG	CGG
AUU	ACU	AAU	AGU
AUC	ACC	AAC	AGC
AUA	ACG	AAA Lys	AGA
AUG	ACA	AAG	AGG
GUU	GCU	GAU	GGU
GUC	GCC	GAC	GGC
GUA	GCA	GAA	GGA
GUG	GCG	GAG	GGG