

ФАКТОРЫ ОТБОРА ПРОТОБИОМОНОМЕРОВ ПРИ ВОЗНИКНОВЕНИИ ГЕНЕТИЧЕСКОГО КОДА *

А.И. Саралов, *Институт экологии и генетики микроорганизмов УрО РАН*

Для цитирования:

Саралов А.И. Факторы отбора протобиономеров при возникновении генетического кода // Вестник Пермского федерального исследовательского центра. – 2021. – № 4. – С. 49–61. <https://doi.org/10.7242/2658-705X/2021.4.5>

Серьезным препятствием при естественном отборе первородных протобиономеров для создания прототипа генетического кода при абиогенезе и в протоклеточную фазу являлось разнообразие структурных, пространственных и поворотных изомеров. Они одинаковы по элементному составу и молекулярной массе (M), но вследствие различного расположения атомов в молекуле могут существенно отличаться по своим физико-химическим свойствам, таким как температура плавления $T_{пл}$, $T_{пл} : M$ и растворимость в воде. Эти параметры различают *цис-транс*-изомеры дикарбоновых кислот, пространственные изомеры моносахаридов и структурные изомеры α , β , γ -аминокислот. Стабильные плоские гетероциклические молекулы главных нуклеотидных оснований состоят из трех (C, H, N) или четырех (C, H, N, O) атомов, содержат по одной $-C=C$ связи и по два атома азота в каждом гетероцикле в $C-N$ и $C=N$ связях. Они существуют как изомеры резонирующих гибридов одинарных и двойных связей и как смесь таутомерных форм из-за наличия $-C=O$ и/или $-NH_2$ боковых групп. Они термостойки, в воде нерастворимы, проявляют признаки стабильности твердого тела, что очень важно для молекул ДНК – носителей наследственной информации. На диаграммах $M-T_{пл}$, протеиногенные аминокислоты и соответствующие им кодоны располагаются довольно упорядоченно относительно обособленных кластеров пуриновых и пиримидиновых оснований. Совокупность сведений об эволюции системы генетического кодирования позволяет заключить, что элементный состав и структура специфичных протобиономеров, их M , $T_{пл}$, фото- и термостабильность, индивидуальная растворимость в воде и гидрофильно-гидрофобные взаимодействия определяли их отбор при зарождении прообраза стандартного генетического кода.

Ключевые слова: абиогенез, генетический код, протобиономеры, свойства аминокислот и нуклеотидных оснований.

* Работа выполнена в рамках государственного задания (регистрационный номер НИОКТРАААА–А–19–119112290008–4).

Введение

Генетический код – свойственная живым организмам единая система «записи» наследственной информации в молекулах нуклеиновых кислот в виде последовательности нуклеотидов [1]. Один триплет нуклеотидов (кодон) кодирует одну аминокислоту, а протяженная последовательность нуклеотидов гена соответствует неперекрывающейся последовательности аминокислот в молекуле синтезируемого белка. Стандартную («каноническую») версию генетического кода обычно представляют в виде таблицы из 64 триплетов для 20 канонических аминокислот и трех стоп-кодонов завершения трансляции, допуская ее «универсальность» почти для всех живущих прокариот и эукариот. Однако в геномах некоторых организмов через стоп-кодоны могут генетически кодироваться дополнительные неканонические аминокислоты: UGA может быть транслирован как селеноцистеин [21], а UAG – как пирролизин [19].

При изучении проблемы происхождения стандартного кода выдвинут ряд взаимодополняющих гипотез.

- Гипотеза «застывшей случайности» рассматривает возможность закрепления раннего кода на основе случайно возникшей привязки кодонов к первым аминокислотам. Новые аминокислоты при адаптивной эволюции соблюдали согласованную связь со старыми, как правило, минимально разрушали последовательности в протеинах, в которые включались [12, 15].

- Теория минимизации ошибок в коде отмечает высокую устойчивость неслучайной структуры стандартного генетического кода к мутациям и ошибкам трансляции [7, 30].

- Гипотеза стерической комплементарности отстаивает важность прямого специфического взаимодействия между аминокислотами и кодирующими их триплетами нуклеотидов – с кодонами и антикодонами [40, 57].

- Теория коэволюции предусматривает, что некоторые протеиногенные аминокислоты были продуцированы рано

при абиогенезе, другие на начальных биогенных стадиях в протоклеточную фазу или после возникновения системы трансляции [14, 56]. Разработана схема, отображающая связи последовательных метаболических превращений простых питательных веществ и аминокислот с соответствующими кодонами [51]. Отмечено два блока синтеза аминокислот. Первый блок связан с фосфорилированием и путями распада глюкозы $C_6H_{12}O_6$, которые ведут к C_3 -соединениям, преимущественно к пирувату $C_3H_4O_3$. Второй блок синтеза аминокислот связан с циклом трикарбоновых кислот с ветвями от оксалоацетата $C_4H_4O_5$ и α -кетоглутарата $C_5H_6O_5$. Первый нуклеотид кодона неизменно связан с путями синтеза аминокислот и с хронологией их появления в коде. Наиболее древние пять аминокислот кодируются кодонами с гуанином в первой позиции (G_1). Процесс биосинтеза аминокислот в первом блоке завершается кодированием кодонами с урацилом в первой позиции (U_1), а во втором блоке – с аденином (A_1).

- Способность РНК сочетать в себе наследственные и каталитические функции позволила выдвинуть плодотворную идею существования «РНК-мира» в ранний период развития жизни до возникновения стандартного генетического кода [55]. В мире РНК в качестве кофакторов олигонуклеотидных рибозимов с различными каталитическими свойствами могли функционировать аминокислоты и пептиды [22, 50]. Этот древний ансамбль кооператоров проторибосомы, по-видимому, обладал способностью катализировать аминокислотное ацилирование, полимеризацию РНК, быть эволюционным предшественником тРНКаз и прообраза системы трансляции [25, 42]. Эти исследования являются частью разрабатываемых различных представлений о происхождении и эволюции генетического кода, системы трансляции и совокупности многих факторов, влияющих на процессы биосинтеза белка на рибосоме [18, 26, 53].

Несмотря на обширную литературу по дискуссионной междисциплинарной проблеме происхождения стандартного генетического кода и его версий, можно констатировать, что значительно слабее изучена физико-химическая сущность естественного отбора специфических протобиомономеров при абиогенезе и в протоклеточную фазу.

Основная цель работы – рассмотреть детерминирующие факторы отбора протомономеров в древнюю фазу самоорганизации прототипа генетического кода.

Образование биоорганических мономеров из простых молекул

В ранних пионерских работах, имитирующих вероятные примордиальные условия в атмосфере древней Земли, удалось экспериментально продемонстрировать возможность абиогенного синтеза некоторых моносахаридов, аминокислот и других мономеров – основных строительных блоков биополимеров [34, 36]. Согласно современным теориям возникновения жизни, молекулы формамида NH_2CHO , которые обладают реактивностью и стабильностью, детектированы в межзвездном пространстве, метеоритах и кометах, могли быть единственным мультифункциональным пребиотическим прекурсором протобиомономеров, нуклеотидов и информационных полимеров, легко получаемых абиотически [4, 41, 43, 46]. Рассматриваются различные сценарии внеземного пребиотического синтеза органических молекул и гомохиральных аминокислот на частицах пыли в регионах образования звезд и на больших метеороидах в пределах магнитных полей формирующихся звезд [9, 17]. В лабораторных экспериментах, имитирующих светящиеся под действием ультрафиолетовой (УФ) и фотоиррадиации холодные газопылевые галактические туманности, получены пиримидиновые и пуриновые азотистые основания [31, 32]. Анализ лимитирующих условий обитания экстремофилов в криосфере Земли позволяет предпола-

гать существование внеземной криожизни в пределах Солнечной системы, прежде всего на Марсе, крупных спутниках Юпитера и Сатурна [2, 29, 33, 47].

В 70-х годах прошлого столетия на гребнях океанических хребтов было обнаружено множество горячих излияний с подводными оазисами жизни. Излияния с температурой 250–370°C, сопровождаемые выбросами полиметаллических сульфидов, известны под названием «черные курильщики». После их описания была предложена ставшая популярной гипотеза возникновения жизни в подобных первобытных гидротермальных системах [35]. По модели Russel и Hall [44], жизнь зарождалась возле горячих излияний в микроскопических везикулоячеистых образованиях, облицованных пластичными минерально-коллоидными полупроницаемыми перегородками с каталитическими сульфидами металлов (FeS , FeS_2 , Fe_3S_4 , NiS , CoS) и органическими сульфидами, но без липопротеидов, пирофосфатов и АТФ, как в протоклетках. Также РНК-мир мог зарождаться при абиогенных геохимических процессах на древней Земле на каталитически активных минеральных поверхностях [54], на гидрофильном глинистом слоистом монтмориллоните [5] или в сети ячеек циано-сульфидного протометаболизма в смеси HCN , HCHO , H_2S , вода, аммоний, полифосфаты, УФ и ионизирующая радиация [37].

В 2000 г. вне Срединно-Атлантического хребта было обнаружено гидротермальное поле, названное «Lost City» [23]. В отличие от вулканогенных «черных курильщиков» эти гидротермальные флюиды менее горячие (150–200°C), сильнощелочные (pH 9–10), богаты водородом, содержат CH_4 , NH_3 и H_2S . Их подводные постройки формируются в процессе серпентинизации при взаимодействии морской воды с железомagneзными силикатными минералами типа оливина [48]. При гидроксилации оливина высвобождаются серпентин, магнетит и H_2 . Именно в подобных щелочных гидротермальных вентах, бо-

гатым водородом, сульфидами железа, и органических соединений с температурой до 200°C могли зародиться каталитические протоцеллюлярные системы [45].

В условиях гидротерм типа «Lost City» из неорганических прекурсоров (CO_2 , H_2 , CH_4 , N_2 , NH_3 и H_2S) предположительно осуществлялся абиогенный синтез аминокислот [24]. В аналогичных восстановительных условиях древних гидротерм из газообразных простых молекул могли синтезироваться активные пребиотические молекулы циановодорода HCN , формальдегида HCHO и формамида NH_2CHO , затем моносахариды из HCHO , пурины из HCN , а из смесей HCN , HCHO или из NH_2CHO – аденин, гуанин, урацил, тимин и цитозин [28, 43, 46].

Факторы отбора протобиономеров

Около 99% массы всего живого построено в основном из четырех легких атомов H , C , N и O , способных образовывать прочные ковалентные связи и активно реагировать друг с другом [38]. Синтез ядер химических элементов связан с энергогенерирующими термоядерными реакциями на массивных звездах при сверхвысоких температурах порядка 10^7 – 10^9 К [8]. Солнечная энергия тепла и света, достигающая поверхности Земли, служит важнейшим внешним фактором существования биосферы. Кислород – продукт фотосинтеза растений в период $0,5$ – $0,3 \times 10^9$ лет тому назад в атмосфере у поверхности Земли достиг 21% (азот 78,1%), составляет половину массы земной коры, 88,8% – массы воды и около 70% – тканей живых существ.

Элементный состав первородных биологических мономеров широко варьирует. Четырехвалентный углерод – важнейший биогенный элемент, является основой для множества молекулярных веществ. Метан CH_4 , первый член ряда неполярных гидрофобных углеводородов, состоит лишь из двух атомов C и H . Вода H_2O тоже состоит лишь из двух атомов O и H , является одним из самых распространенных веществ на Земле, без которого возникновение и существование жизни

невозможно. Наличие полярной жидкой воды, процессы гидратации и гидрофильно-гидрофобные взаимодействия являются важными факторами химической и биологической эволюции [27].

Обсуждаются возможности существования криожизни в условиях без жидкой воды и O_2 , например на Титане – наиболее крупном спутнике Сатурна [11, 33]. На его поверхности детектирована криогенная смесь $\text{N}_2 + \text{CH}_4 + \text{C}_2\text{H}_6$. Неполярные углеводородные природные жидкости на основе CH_4 и C_2H_6 , возможно, служат альтернативными растворителями для аборигенной криожизни на Титане. Экспериментально показано формирование не липосом, а нового типа эластичных гидрофобных клеточных мембран – азотосом, которые могут создаваться и функционировать в жидком метане при криогенных температурах [49].

Если в молекулы кроме C и H дополнительно входит кислород, то резко возрастает разнообразие химических соединений и их изомеров, различающихся по строению и, следовательно, по физико-химическим свойствам (таблица). Углеводы $(\text{CH}_2\text{O})_n$, содержащие одну карбонильную $-\text{C}=\text{O}$ и несколько гидроксильных $-\text{OH}$ групп, характеризуются разнообразием асимметричных пространственных стереоизомеров, но лишь некоторые из них отобраны для создания жизнеспособных систем (глюкоза, рибоза и дезоксирибоза).

Богаты пространственными и структурными изомерами карбоновые кислоты, состоящие тоже лишь из C , H , O и содержащие одну или несколько трехатомных карбоксильных $-\text{COOH}$ групп. При переходе от монокарбоновых к дикарбоновым кислотам резко повышается $T_{\text{пл}}$ и отношение $T_{\text{пл}} : M$, но утрачивается неограниченная растворимость в воде. Существенно могут различаться физико-химические свойства у *цис-транс*-стереоизомеров дикарбоновых кислот. Например, у более гидрофобного вытянутого фумарового *транс*-изомера (по сравнению с изогнутым малеиновым *цис*-изомером) в 113 раз ниже растворимость в H_2O , но в 2,2 раза выше $T_{\text{пл}}$ и $T_{\text{пл}} : M$.

Физико-химические свойства основных и минорных протобиомономеров

Соединение и формула	М	$T_{пл}$, °С	$T_{пл}/M$	Растворимость в H_2O (25°С)
Пировиноградная кислота $C_3H_4O_3$	88,1	13,6	0,15	∞
Малеиновая кислота $C_4H_4O_4$	116,1	140	1,20	79
Фумаровая кислота $C_4H_4O_4$	116,1	302	2,60	0,7
Дезокси-D-рибоза $C_5H_{10}O_4$	134,1	98	0,73	Х. р.
D-Рибоза $C_5H_{10}O_5$	150,1	87	0,58	Х. р.
D-Арабиноза $C_5H_{10}O_5$	150,1	160	1,07	59
β -D-Фруктоза $C_6H_{12}O_6$	180,2	105	0,58	375
β -D-Галактоза $C_6H_{12}O_6$	180,2	167	0,93	68
β -Аланин $H_2N(CH_2)_2COOH$	89,1	207	2,32	54,5
α -Аланин $CH_3CH(NH_2)COOH$	89,1	314	3,52	16,7
L-Норлейцин $CH_3(CH_2)_3CH(NH_2)COOH$	131,2	301	2,29	1,5
L-Лейцин $\begin{matrix} CH_3 \\ \\ CH \\ \\ CH_3 \end{matrix} CH_2CH(NH_2)COOH$	131,2	295	2,25	2,4
L-Изолейцин $CH_3CH_2CH \begin{matrix} CH_3 \\ \\ CH \end{matrix} (NH_2)COOH$	131,2	285	2,17	4,1
Цитозин $C_4H_5N_3O$	111,1	224	2,02	0,77
5-Азацитозин $C_3H_4N_4O$	112,1	300	2,68	0,004
5-Флуороцитозин $C_4H_4N_3OF$	129,1	297	2,30	150
Урацил $C_4H_4N_2O_2$	112,1	283	2,52	0,36
5-Метилцитозин $C_5H_7N_3O$ (5-mC)	125,1	221	1,77	0,45
Тимин (5-Метилурацил) $C_5H_6N_2O_2$	126,1	282	2,24	0,40
2-Аминопурин $C_5H_5N_5$	135,1	282	2,09	0,12
Аденин $C_5H_5N_5$ (6-Аминопурин)	135,1	305	2,28	0,09
N6-Метиладенин $C_6H_7N_5$ (N6-mA)	149,2	300	2,02	1,18
Гипоксантин $C_5H_4N_4O$ (Нур)	136,1	310	2,28	0,09
Гуанин $C_5H_5N_5O$	151,1	346	2,29	0,004
7-Метилгуанин $C_6H_7N_5O$ (7-mG)	165,2	370	2,24	$\leq 0,004$
6-Тиогуанин $C_5H_5N_5S$	167,2	314	1,92	22,5

Богатые атомами кислорода углеводы и жирные кислоты хорошо растворимы в воде, склонны к образованию сравнительно легкоплавких кристаллогидратов. Так, безводные кристаллы D-глюкозы из метанола имеют $T_{пл}$ 146°С, а у моногидрата из H_2O $T_{пл}$ ниже на 63°С [13].

Если в органические соединения с атомами С, Н, О дополнительно включаются атомы азота, то они могут приобретать повышенную плотность, $T_{пл}$ и $T_{кип}$, но пониженную растворимость в H_2O , что позволяет им значительно легче кристаллизоваться. Например, превращение пирувата в аланин путем замещения атома кислорода на NH_3 сопровождается

повышением М лишь на 1, но резким увеличением $T_{пл}$ с 14 до 314°С ($T_{пл} : M$ с 0,15 до 3,52) и утратой неограниченной растворимости в воде. Также при замещении у формальдегида $HCHO$ одного атома водорода аминогруппой у производного формамида NH_2CHO резко повышается $T_{кип}$ на 230,2°С, $T_{пл}$ на 94,6°С, разность $T_{кип} - T_{пл}$ на 135°С и плотность на 0,22 г/см³. Формамид обладает химической реактивностью и стабильностью, обнаружен в межзвездном пространстве, полагают, пригоден в качестве пребиотического прекурсора в химической эволюции и возникновении жизни на Земле и других планетах [4, 46].

Особого внимания заслуживает тот факт, что у структурных изомеров α , β , γ -аминокислот могут существенно изменяться $T_{пл}$ и растворимость в воде. Напротив, у структурных изомеров α -аминокислот, различающихся порядком расположения углеводородных групп $-\text{CH}_2$ и $-\text{CH}_3$, отличия в растворимости, $T_{пл}$ и $T_{пл} : M$ незначительны. В свете этих данных, представляется вполне оправданным предположение, что норлейцин и норвалин могли быть довольно обычными компонентами протеинов на ранних стадиях клеточной эволюции [6].

Главные азотистые основания состоят из трех (C, H, N) или четырех (C, H, N, O) атомов, содержат по одной двойной $-\text{C}=\text{C}$ связи и по 2 атома азота в каждом гетероцикле в составе одинарных или двойных связей (C–N, C=N). Стабильные плоские циклические молекулы азотистых оснований существуют в виде изомерных резонансных гибридов одинарных и двойных связей и одновременно из смеси двух и более таутомерных форм благодаря наличию боковых $-\text{C}=\text{O}$ и/или $-\text{NH}_2$ групп [3]. Среди них нет поворотных изомеров (конформеров) с насыщенной углеродной связью, придающей им неплоскую конформацию, снижающую стабильность структуры нуклеиновых кислот. Попытки модифицировать главные и минорные основания, например, путем замещения в гетероцикле одного из атомов углерода на атом азота или путем дополнительного введения в боковые группы атомов фтора или серы, сопровождаются резкими негативными изменениями $T_{пл}$ и особенно – растворимости в H_2O .

Главные гетероциклические основания характеризуются термостабильностью и низкой растворимостью в воде. Наименее стабилен среди них цитозин: сравнительно низка $T_{пл}$, лучше растворим в воде, легко дезаминируется до урацила, не обнаружен в метеоритах [39]. Гидрофобность главных азотистых оснований тесно связана с явлением резонанса [38]. Наибольшей энергией резонанса среди них обладает стабильная бескислородная молекула аденина $\text{C}_5\text{H}_5\text{N}_5$

(6-аминопурин). Столь высокой стабильности, как у аденина, нет у его структурного изомера 2-аминопурина, обладающего способностью к интенсивной флуоресценции. Эволюционный отбор на фотостабильность, обусловленную коротким временем жизни их электронного возбуждения светом, вероятно, прошли при абиогенезе и другие природные азотистые основания нуклеиновых кислот.

При воссоздании ранних стадий эволюции триплетного кода подчеркивалась важность абиогенного старта [16]. Первые 5 аминокислот (Gly, Ala, Asp, Glu, Val) могли иметь повышенную исходную концентрацию в локальном «первичном бульоне». Именно эти 5 аминокислот из 10 первоначальных обнаружены в метеоритах [10], синтезированы в лабораторных экспериментах, имитирующих примордиальные условия в атмосфере древней Земли и возле морских гидротерм из CO_2 , H_2 , CH_4 , NH_3 и H_2S [24, 36].

Trifonov [52] на основе комплексного анализа 60 физических, химических и биохимических свойств аминокислот и соответствующих им кодонов/антикодонов установил временную последовательность («хронологию») аминокислот и кодонов в генетическом коде. Финальный консенсус выявил следующий временной порядок канонических аминокислот: 1-Gly, 2-Ala, 3-Asp, 4-Val, 5-Pro, 6-Ser, 7-Glu, 8-Thr, 9-Leu, 10-Arg, 11-Ile, 12-Gln, 13-Asn, 14-His, 15-Lys, 16-Cys, 17-Phe, 18-Tyr, 19-Met, 20-Trp. Затем Higgs и Pudritz [20] для углубленного анализа хронологии аминокислот в коде использовали дополнительные термодинамические критерии. Установленный ими временной порядок аминокислот почти полностью совпал с таковым по Trifonov [52], но Pro был переставлен с пятой на седьмую позицию вместо Glu. Также, согласно теории коэволюции [51], Glu напрямую продуцируется из α -кетоглутарата, а Pro является производной иминокислотой из Glu. Поэтому логически следует внести коррективы в хронологию этих двух аминокислот по Trifonov.

Влияние М и $T_{пл}$ на взаимосвязь аминокислот и кодонов

Мы впервые показали, что на диаграмме $M-T_{пл}$ главные и минорные нуклеотидные основания группируются в два четко обособленных кластера – классы пуриновых и пиримидиновых гетероциклов (рис. 1). У главных оснований при незначительном повышении М существенно увеличивается их $T_{пл}$ и отношение $T_{пл} : M$. Однако увеличение М за счет метилирования главных оснований обычно не сопровождается повышением $T_{пл}$. Было установлено, что аминокислоты и соответствующие им кодоны на диаграмме $M-T_{пл}$ располагаются довольно системно относительно кластеров пиримидиновых и пуриновых оснований. Например, пять аминокислот, кодируемых кодоном с урацилом во второй позиции (U_{II}), располагаются компактно в узком диапазоне М и $T_{пл}$ между кластерами пуриновых и пиримидиновых оснований.

Здесь U_{II} кодоны отличаются один от другого степенью вырожденности (1–2–3–4–6), а кодируемые ими гидрофобные аминокислоты – хронологической последовательностью появления в генетическом коде (# 4–9–12–17–19). Напротив, шесть аминокислот (# 6 Ser, 9 Leu, 16 Cys, 17 Phe, 1 Tyr, 20 Trp), кодируемых кодоном с урацилом, в первой позиции (U_I) на диаграмме располагаются в широком диапазоне М и $T_{пл}$. Три аминокислоты (# 9–17–20), кодируемые кодоном UU_{AG} , UU^{UC} и UGG , существенно отличаются по М (131,2–204,2), но имеют сходные $T_{пл}$ (285–295°C). В конце последовательности после двукратных кодонов UA^{UC} для наиболее термостойкого Tyr следуют стоп-кодоны UA_{AG} . Также после однократного кодона UGG для наиболее массивного Trp следует стоп-кодон UGA . Через стоп-кодон иногда ферментативно кодируется # 20 Trp или # 21 SeC, а через UAG – # 22 Pyl [19, 21].

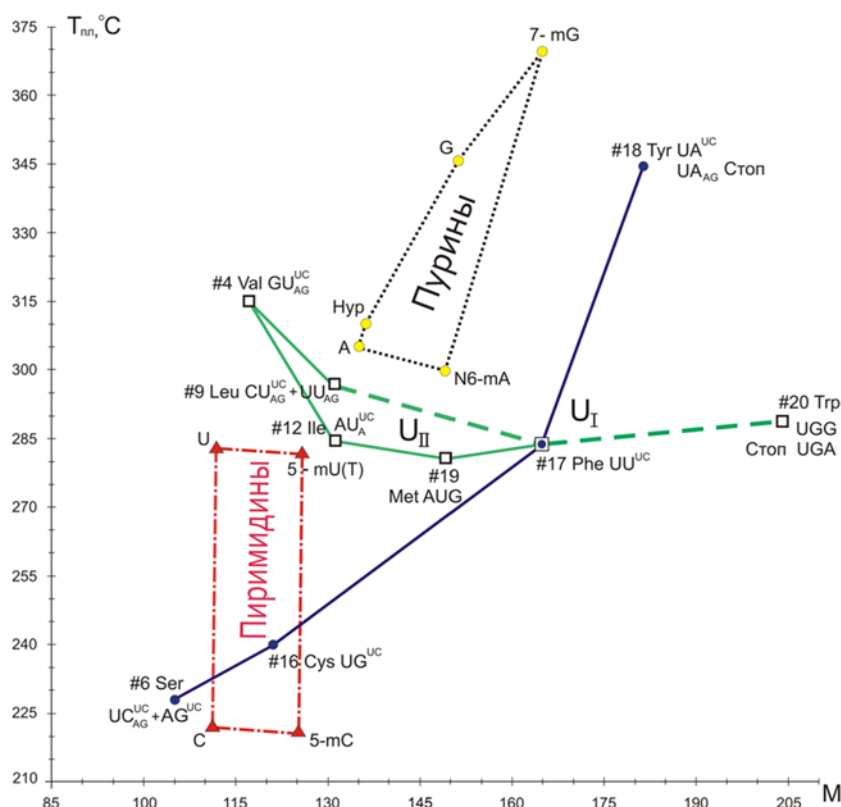


Рис. 1. Значения М и $T_{пл}$ пуриновых и пиримидиновых оснований и канонических аминокислот, кодируемых кодоном с урацилом в первом (U_I) или втором (U_{II}) положении

Обозначения кодоновых оснований: Y – пиримидины; R – пурины; N – любое; $NN^Y = NN^{UC}$; $NN_R = NN_{AG}$; $NN_R^Y = NN_{AG}^{UC}$; #, хронологический порядок аминокислот в генетическом коде

На следующей диаграмме (рис. 2) триплеты с цитозином в первой (C_I) или второй (C_{II}) позициях кодируют по четыре аминокислоты, которые в каждом кластере сходны по M , но сильно отличаются по $T_{пл}$ и $T_{пл} : M$. Аминокислоты (# 2–6–7–8), кодируемые четырех- и шестикратными кодонами с C_{II} и N_I ($G-U-C-A$), сравнительно низкомолекулярны, но достаточно термостойки, являются нейтральными, часто встречаются в белках, где обычно выступают в роли простейших связующих звеньев. Эта группа аминокислот включает иминокислоту # 7 Pro, хотя кодируется четырехкратными кодонами с C_I и C_{II} . Вероятно, действительно вторая позиция играла ключевую роль в образовании триплетов и их соответствия с подобными аминокислотами [16]. Более молодые аминокислоты (# 9–10–13–14), кодируемые шести- и двукратными кодонами с C_I , более высокомолекулярны с пониженным отношением $T_{пл} : M$, проявляют нейтральные или основные свойства (His, Arg). Следовательно, при адаптив-

ной эволюции раннего кода новые аминокислоты соблюдали согласованную связь с предшествующими аминокислотами, группируясь по сходным физико-химическим признакам [15, 20].

Третья диаграмма $M-T_{пл}$ (рис. 3) отражает соответствие между кодонами с G_I и A_{II} и аминокислотами, которое утанавливалось в разные исторические периоды формирования генетического кода. Здесь кластер кодонов с G_I кодирует пять наиболее древних C_2-C_5 аминокислот (# 1–5). Первые три короткоцепочечные, но сравнительно термостойкие неполярные аминокислоты (Gly, Ala, Val с $T_{пл} : M$ 3,5–2,7) кодируются синонимическими квартетами кодонов. Менее термостойкие дикарбоновые полярные аминокислоты (Asp, Glu с $T_{пл} : M$ 2,0–1,7) кодируются двукратными кодонами с G_I и одновременно входят в кластер кодонов с A_{II} . Семь аминокислот, кодируемые двукратными кодонами с аденином во второй позиции (A_{II}), характеризуются M от среднего до

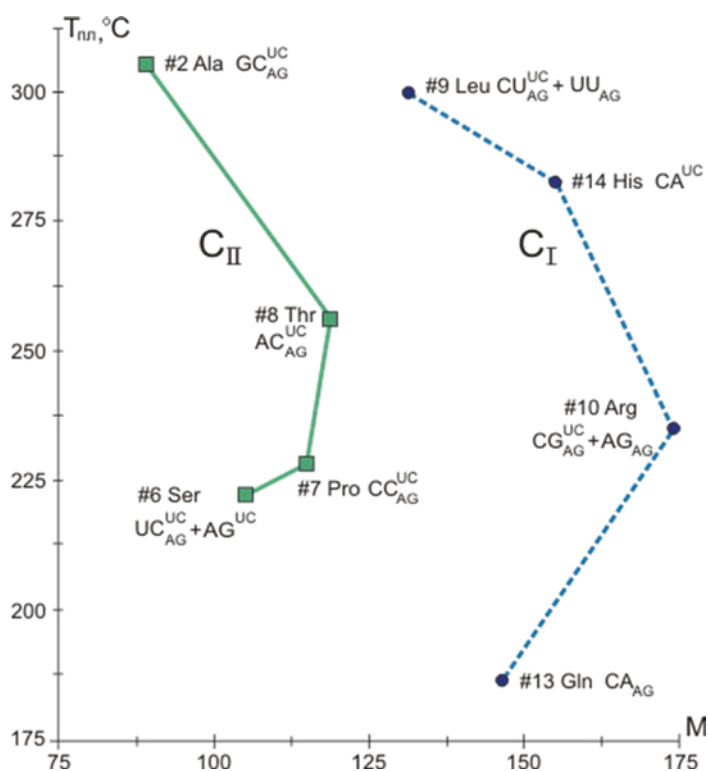


Рис.2. Распределение канонических аминокислот, кодируемых кодонами с цитозином в первом (C_I) или втором (C_{II}) положении в зависимости от M и $T_{пл}$

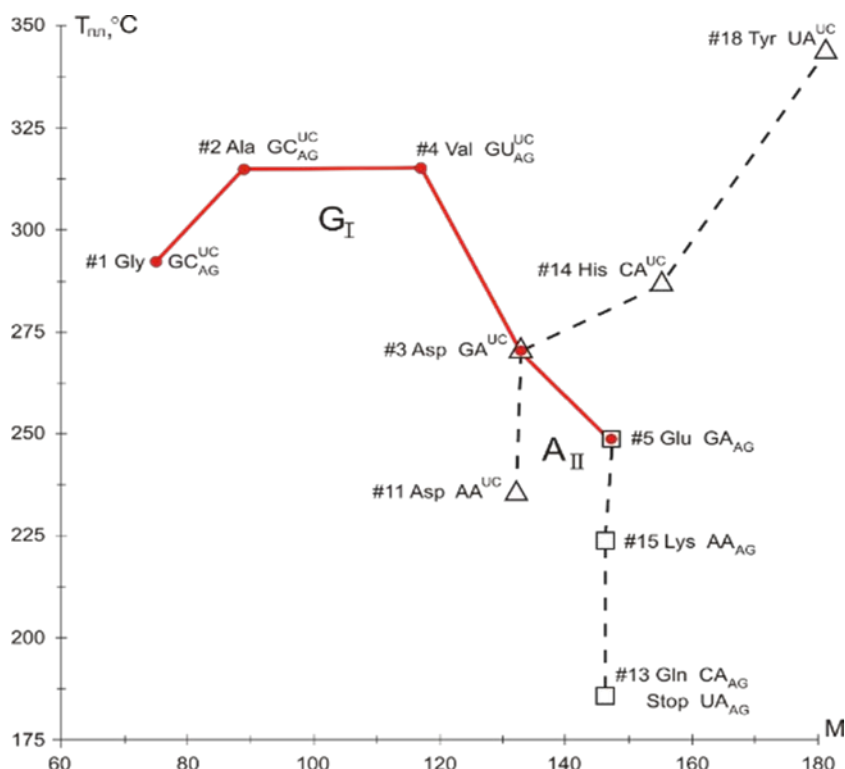


Рис. 3. Распределение канонических аминокислот, кодируемых кодами с гуанином в первом (G_I) или аденином во втором (A_{II}) положении в зависимости от M и T_{mI}

высокого значения (132–181) и располагаются в рекордно широком диапазоне T_{mI} от 186°C у Gln до 344°C у Tyr. Эти аминокислоты в зависимости от содержания в третьей позиции кодонов пиримидинов (NA^{UC}) или пуринов (NA_{AG}) принадлежат разным ветвям. Одна ветвь (обозначена $\triangle---\triangle$) сравнительно более термостойких аминокислот с T_{mI} 236–344°C кодируется дублетами кодонов с пиримидинами в третьей позиции NA^{UC} . Здесь последовательность оснований в первой позиции $N = A, G, C, U$ отражает изменения от AA^{UC} (для Asp) до UA^{UC} (для наиболее термостойкого Tyr). Другая ветвь (обозначена $\square---\square$) включает менее термостойкие аминокислоты (T_{mI} 249–186°C), но практически сходные M 146,2–147,1. Они кодируются кодами с пуринами в третьей позиции NA_{AG} . Здесь последовательность оснований в первой позиции $N = G, A, C, U$ завершается стоп-кодонами UA_{AG} (после CA_{AG} для самого легкоплавкого Gln).

Есть предпосылки полагать, что двукратно и однократно вырожденные ко-

доны с U_I фактически завершали формирование структуры стандартного генетического кода, включая кодирование # 16–18, # 20–22 аминокислот и стоп-кодоны (UA_{AG} и UGA). Следовательно, в конечном итоге при установлении соответствия между аминокислотами и кодами из их состава был полностью исключен термодинамически менее стабильный, сравнительно легкоплавкий и наиболее растворимый в воде цитозин. Напротив, первые четыре кодона могли формироваться по шаблону GNC [16].

Заключение

Исходный синтез и естественный отбор специфичных протобиомонеров при воссоздании прообраза стандартного генетического кода при абиогенезе и в протоклеточную фазу, по-видимому, осуществлялся при согласованном сочетании совокупности ряда детерминирующих факторов.

- Взаимосвязь материи, энергии и наследственной информации при развитии органического мира от простого к

сложному. Масса – одна из основных физических характеристик материи, определяющая ее инертные и гравитационные свойства. Вода – одно из наиболее распространенных веществ на Земле, без которого зарождение и существование жизни невозможно. Энергия – общая мера движения и взаимодействия материи; температура – производная от кинетической энергии атомов и молекул вещества. РНК, обладающая способностью к катализу и хранению наследственной информации, вероятно, играла ключевую роль при возникновении РНК-мира и жизни.

- Сопряженность геофизических и геохимических процессов на первобытной Земле (при наличии каталитических сульфидов металлов в везикулярических комплексах, ультрафиолетовой и ионизирующей радиации) с абиогенезом во Вселенной. Солнечная энергия тепла и света является важнейшим внешним фактором существования современной жизни на Земле.

- Наличие у протобиомономеров отличительных физико-химических свойств и признаков: небольшая молекулярная масса, своеобразие элементно-

го состава и структуры молекул, фото- и термостабильность, индивидуальная растворимость в воде, процессы гидратации и гидрофильно-гидрофобные взаимодействия, уникальность двадцати канонических аминокислот и пяти природных азотистых оснований РНК и ДНК.

- Установлено, что на диаграмме в координатах $M-T_{пл}$ канонические аминокислоты вместе с кодирующими кодонами располагаются вполне закономерно относительно обособленных групп пуриновых и пиримидиновых оснований, отражая соответствие между ними, которое устанавливалось в определенные периоды формирования генетического кода.

- Показано, что существенное препятствие при отборе первородных протобиомономеров представляло многообразие структурных, пространственных и поворотных изомеров. Они сходны по элементному составу и молекулярной массе, но вследствие различного расположения атомов могут значительно отличаться по своим физико-химическим свойствам, в частности, по $T_{пл}$, $T_{пл} : M$ и растворимости в воде.

Библиографический список

1. Анпель Б., Бенке Б.-И., Бененсон Я. [и др.] Нуклеиновые кислоты: От А до Я – М.: БИНОМ. Лаборатория знаний, 2015. – 413 с.
2. Саралов А.И. Адаптивность экстремофилов *Archaea* и *Bacteria* // Микробиология. – 2019. – Т. 88. – № 4. – С. 377–400.
3. Саралов А.И. Расчет энтальпий сгорания аминокислот и азотистых оснований по их элементному составу и структуре // Международный журнал прикладных и фундаментальных исследований. – 2020. – № 3. – С. 113–121.
4. Adande G.R., Woolf N.J., Ziurys L.M. Observations of interstellar formamide: availability of a prebiotic precursor in the galactic habitable zone // *Astrobiology*. – 2013. – Vol. 13. – № 5. – P. 439–453.
5. Aldersley M.F., Joshi P.C., Price J.D., Ferris J.P. The role of montmorillonite in its catalysis of RNA synthesis // *Applied Clay Science*. – 2011. – Vol. 54. – № 1. – P. 1–14.
6. Alvarez-Carreño C., Bacerra A., Lazcano A. Norvaline and norleucine may have been more abundant protein components during early stages of cell evolution // *Orig. Life. Evol. Biosph.* – 2013. – Vol. 43. – № 4–5. – P. 363–375.
7. Ardell D.H. On error minimization in sequential origin of the standard genetic code // *J. Mol. Evol.* – 1998. – Vol. 47. – № 1. – P. 1–13.
8. Bethe H.A. Energy production in Stars // *Science*. – 1968. – Vol. 161. – № 3841. – P. 541–547.
9. Boyd R.N., Kajino T., Onaka T. Supernovae, neutrinos and the chirality of amino acids // *Int. J. Mol. Sci.* – 2011. – Vol. 12. – № 6. – P. 3432–3444.
10. Burton A.S., Stern J.C., Elsila J.E., Glavin D.P., Dworkin J.P. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites // *Chem. Soc. Rev.* – 2012. – Vol. 41. – № 16. – P. 5459–5472.

11. *Cordier D., Garcia-Sánchez F., Justo-Garcia D.N., Liger-Belair G.* Bubble streams in Titan's seas as a product of liquid N₂ + CH₄ + C₂H₆ cryogenic mixture // *Nature Astronomy*. – 2017. – 1, 0102. – P. 1–4.
12. *Crick F.H.C.* The origin of the genetic code // *J. Mol. Biol.* – 1968. – Vol. 38. – № 3. – P. 367–379.
13. *Dawson R.M.C., Elliot D.C., Elliot W.H., Jones K.M.* Data for biochemical research – Oxford: Clarendon press, 1986. – 580 с.
14. *Di Giulio M.* Some pungent arguments against the physico-chemical theories of the origin of the genetic code and corroborating the coevolution theory // *J. Theor. Biol.* – 2017. – Vol. 414. – P. 1–4.
15. *Doig A.J.* Frozen, but no accident—why the 20 standard amino acids were selected // *FEBS J.* – 2017. – Vol. 284. – № 9. – P. 1296–1305.
16. *Eigen M., Schuster P.* The Hypercycle: A principle of natural self-organization – Springer-Verlag, Berlin, Heidelberg, New York, 1979.
17. *Famiano M.A., Boyd R.N., Kajino T., Onaka T., Mo Y.* Amino acid chiral selection via weak interactions in stellar environments: Implications for the origin of life // *Sci. Rep.* – 2018. – 8, 8033.
18. *Francis B.R.* Evolution of the genetic code by incorporation of amino acids that improved or changed protein function // *J. Mol. Evol.* – 2013. – Vol. 77. – № 4. – P. 134–158.
19. *Gaston M.A., Zhang L., Green-Church K.B., Krzycki J.A.* The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine // *Nature*. – 2011. – Vol. 471. – № 7340. – P. 647–650.
20. *Higgs P.G., Pudritz R.E.* A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code // *Astrobiology*. – 2009. – Vol. 9. – № 5. – P. 483–490.
21. *Hohn M.J., Palioura S., Su D., Yuan J., Söll D.* Genetic analysis of selenocysteine biosynthesis in the archaeon *Methanococcus maripaludis* // *Mol. Microbiol.* – 2011. – Vol. 81. – № 1. – P. 249–258.
22. *Huber C., Eisenreich W., Hecht S., Wächtershäuser G.* A possible primordial peptide cycle // *Science*. – 2003. – Vol. 301. – № 5635. – P. 938–940.
23. *Kelley D.S., Karson J.A., Blackman D.K., Früh-Green G.L., Butterfield D.A., Lilley M.D., Rivizzigno P.* An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30°N // *Nature*. – 2001. – Vol. 412. – № 6843. – P. 145–149.
24. *Kitadai N.* Energetics of amino acid synthesis in alkaline hydrothermal environments // *Orig. Life Evol. Biosph.* – 2015. – Vol. 45. – № 4. – P. 377–409.
25. *Knight R.D., Landweber L.E.* The early evolution of the genetic code // *Cell*. – 2000. – Vol. 101. – № 6. – P. 569–572.
26. *Koonin E.V., Novozhilov A.S.* Origin and evolution of the universal genetic code // *Annu. Rev. Genet.* – 2017. – Vol. 51. – № 1. – P. 45–62.
27. *Krishnamurthy R.* Role of pKa of nucleobases in the origins of chemical evolution // *Acc. Chem. Res.* – 2012. – Vol. 45. – № 12. – P. 2035–2044.
28. *LaRowe D.E., Regnier P.* Thermodynamic potential for the abiotic synthesis of adenine, cytosine, guanine, thymine, uracil, ribose, and deoxyribose in hydrothermal systems // *Orig. Life Evol. Biosph.* – 2008. – Vol. 38. – № 5. – P. 383–397.
29. *Martin A., McMinn A.* Sea ice, extremophiles and life on extra-terrestrial ocean worlds // *Int. J. Astrobiol.* – 2018. – Vol. 17. – № 1. – P. 1–16.
30. *Massey S.E.* The neutral emergence of error minimized genetic codes superior to the standard genetic code // *J. Theor. Biol.* – 2016. – Vol. 408. – P. 237–242.
31. *Materese C.K., Nuevo M., Bera P.P., Lee T.J., Sandford S.A.* Thymine and other prebiotic molecules produced from the ultraviolet photo-irradiation of pyrimidine in simple astrophysical ice analogs // *Astrobiology*. – 2013. – Vol. 13. – № 10. – P. 948–962.
32. *Materese C.K., Nuevo M., Sandford S.A.* The formation of nucleobases from the ultraviolet photo-irradiation of purine in simple astrophysical ice analogues // *Astrobiology*. – 2017. – Vol. 17. – № 8. – P. 761–770.
33. *McKay C.P.* Titan as the Abode of Life // *Life*. – 2016. – Vol. 6. – № 8. – P. 1–15.
34. *Miller S.L.* Production of some organic compounds under possible primitive earth conditions // *J. Am. Chem. Soc.* – 1953. – Vol. 77. – № 9. – P. 2351–2361.
35. *Miller S.L., Bada J.L.* Submarine hot springs and the origin of life // *Nature*. – 1988. – Vol. 334. – № 6183. – P. 609–611.
36. *Miller S.L., Orgel L.E.* The Origins of Life on the Earth – Prentice-Hall, Englewood Cliffs, New Jersey, 1974. – 229 p.
37. *Patel B.H., Percivalle C., Ritson D.J., Duffy C.D., Sutherland J.D.* Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism // *Nat. Chem.* – 2015. – Vol. 7. – № 4. – P. 301–307.
38. *Pauling L., Pauling P.* Chemistry – San Francisco: Freeman W.H. and Company, 1975. – 792 p.
39. *Pearce B.K.D., Pudritz R.E.* Seeding the pregenetic Earth: Meteoritic abundances of nucleobases and potential reaction pathways // *Astrophys. J.* – 2015. – Vol. 807. – № 85. – P. 1–10.

40. *Pelc S.R., Welton M.G.E.* Stereochemical relation between coding triplets and amino acids // *Nature*. – 1966. – Vol. 209. – № 5026. – P. 868–870.
41. *Pino S., Sponer J.E., Costanzo G., Saladino R., Di Mauro E.* From formamide to RNA the path is tenuous but continuous // *Life*. – 2015. – Vol. 5. – № 1. – P. 372–384.
42. *Rodin A.S., Szathmary E., Rodin S.N.* On origin of genetic code and tRNA before translation // *Biol. Direct*. – 2011. – Vol. 6. – № 14. – P. 1–24.
43. *Ruiz-Mirazo K., Briones C., de la Escosura A.* Prebiotic systems chemistry: new perspectives for the origins of life // *Chem. Rev.* – 2014. – Vol. 114. – № 1. – P. 285–366.
44. *Russell M.J., Hall A.J.* The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front // *J. Geol. Soc. London*. – 1997. – Vol. 154. – № 3. – P. 377–402.
45. *Russell M.J., Hall H.J., Martin W.* Serpentinization as a source of energy at the origin of life // *Geobiology*. – 2010. – Vol. 8. – № 5. – P. 355–371.
46. *Saladino R., Crestini C., Pino S., Costanzo G., Di Mauro E.* Formamide and the origin of life // *Phys. Life Rev.* – 2012. – Vol. 9. – № 1. – P. 84–104.
47. *Schulze-Makuch D., Airo A., Schirmack J.* The adaptability of life on Earth and the diversity of planetary habitats // *Front. Microbiol.* – 2017. – Vol. 8. – № 2011. – P. 1–10.
48. *Sleep N.H., Meibom A., Fridriksson T., Coleman R.G., Bird D.K.* H₂-rich fluids from serpentinization: geochemical and biotic implications // *Proc. Natl. Acad. Sci. USA*. – 2004. – Vol. 101. – № 35. – P. 12818–12823.
49. *Stevenson J., Lunne J., Clancy P.* Membrane alternatives in worlds without oxygen: Creation of an azotosome // *Astrobiology*. – 2015. – 1:e1400067. – P. 1–8.
50. *Szathmary E.* Coding coenzyme handles: a hypothesis for the origin of the genetic code // *Proc. Natl. Acad. Sci. USA*. – 1993. – Vol. 90. – № 21. – P. 9916–9920.
51. *Taylor F.J.R., Coates D.* The code within the codons // *Biosystems*. – 1989. – Vol. 22. – № 3. – P. 177–187.
52. *Trifonov E.N.* The triplet code from first principles // *J. Biomol. Struct. Dyn.* – 2004. – Vol. 22. – № 1. – P. 1–11.
53. *Tupper A.S., Pudritz R.E., Higgs P.G.* Can the RNA World still function without cytidine? // *Mol. Biol. Evol.* – 2020. – Vol. 37. – № 1. – P. 71–83.
54. *Wächtershäuser G.* Before enzymes and templates: theory of surface metabolism // *Microbiol. Rec.* – 1988. – Vol. 52. – № 4. – P. 452–484.
55. *Woese C.R.* *The Genetic Code: the molecular basis for genetic expression* – New York: Harper and Row, 1967.
56. *Wong J.T.-F.* A co-evolution theory of the genetic code // *Proc. Natl. Acad. Sci. USA*. – 1975. – Vol. 72. – № 5. – P. 1909–1912.
57. *Yarus M.* The genetic code and RNA-amino acid affinities // *Life*. – 2017. – Vol. 7. – № 13. – P. 1–16.

FACTORS OF PROTOBIOMONOMER SELECTION FOR THE EMERGENCE THE GENETIC CODE

A.I. Saralov

Institute of Ecology and Genetics of Microorganisms UB RAS

For citation:

Saralov A.I. Factors of protobionomer selection for the emergence the genetic code // *Perm Federal Research Center Journal*. – 2021. – № 4. – P. 49–61. <https://doi.org/10.7242/2658-705X/2021.4.5>

The natural selection of specific protobionomers during abiogenic development of the prototype genetic code is hindered by the diversity of structural, spatial, and rotational isomers. They have identical elemental composition and molecular mass (M), but can vary significantly in their physicochemical characteristics, such as melting point T_m, the T_m : M ratio, and solubility in water, due to the different arrangement of atoms in the molecule. These parameters differ between *cis*- and *trans*-isomers of dicarboxylic acids, spatial monosaccharide isomers, and structural isomers of α-, β-, and γ-amino acids. The stable planar heterocyclic molecules of the major nucleobases comprise four (C, H, N, O) or three (C, H, N) elements and contain a single –C=C bond and two nitrogen atoms in each heterocycle involved in C–N and C=N bonds. They exist as isomeric resonance hybrids of

single and double bonds and as a mixture of tautomer forms due to the presence of $-C=O$ and/or $-NH_2$ side groups. They are thermostable, insoluble in water, and exhibit solid-state stability, which is of central importance for DNA molecules as carriers of genetic information. In $M-T_m$ diagrams, proteinogenic amino acids and the corresponding codons are distributed in a rather ordered manner relative to the distinct clusters of purine and pyrimidine bases, reflecting the correspondence between codons and amino acids that was established in different periods of genetic code development. The body of data on the evolution of the genetic coding system indicates that the elemental composition and molecular structure of protobiomonomers, and their M , T_m , photostability, and aqueous solubility determined their selection in the emergence of the standard genetic code.

Keywords: genetic code, origin, chemical evolution, natural selection, properties of amino acids and nucleobases.

Сведения об авторе

Саралов Александр Иванович, доктор биологических наук, профессор, заведующий лабораторией водной микробиологии, Институт экологии и генетики микроорганизмов УрО РАН – филиал Пермского федерального исследовательского центра УрО РАН («ИЭГМ УрО РАН»), 614081, г. Пермь, ул. Голева, 13; e-mail: saralov@iegm.ru

Материал поступил в редакцию 29.06.2021 г.