

Изучение эволюции глицин-аргинин насыщенных доменов низкой сложности

Научный руководитель – **Алексеевский Андрей Владимирович**

Котюргин Александр Павлович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: alekoksan@gmail.com

Длинные участки белков, с превалированием малого числа аминокислот (LCR — low complexity region) широко представлены в белках различных функций, чаще всего — регуляторных, белки с LCR чаще участвуют в белок-белковых взаимодействиях [2, 5]. Для LCR характерны высокая изменчивость и структурное разнообразие [3].

Эта работа началась с изучения эволюции глицин-аргинин насыщенных (GAR) доменов фибрилларинов. Было показано, что классические методы филогении не пригодны из-за неоднозначности при построении выравниваний доменов низкой сложности, но по специфическим признакам можно определить таксономическую принадлежность на уровне отрядов.

С помощью созданной программы в базе данных Uniref100 были найдены 141608 GAR доменов. Среди них обнаружены все 925 GAR доменов фибрилларинов. Критерий домена — участок белка длиной 28 остатков и более с частотой глицинов выше 50% и аргининов выше 10%.

Для всех GAR-содержащих белков описаны домены Pfam, термины GO, найдены содержащие их кластеры гомологичных белков. Показано, что GAR-содержащие белки преимущественно взаимодействуют с РНК, что соответствует ранее опубликованным данным на меньшем материале [1, 4, 6, 7], наиболее частая активность — хеликазная. Помимо этого GAR-домены встречаются в белках с мотивом «цинковый палец», белках, интегрированных в мембрану и убиквитин-ассоциированных белках.

В кластерах ортологичных белков с наибольшей представленностью GAR-домена он присутствует не во всех последовательностях, что позволяет изучить эволюцию домена от его появления. Описано распределение GAR-доменов по надцарствам: наименьшее содержание GAR-доменов в белках архей, белки вирусов же почти целиком относятся к дцДНК-вирусам без РНК-стадии. На примере фибрилларина показано, что GAR-домен медленно меняет заряд в ходе эволюции: 75% доменов эукариот имеют формальный заряд в диапазоне от +16 до +18.

Планируется в дальнейшем найти формальные критерии, позволяющие восстановить филогению по последовательности GAR-домена.

Источники и литература

- 1) Birney E. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors // *Nucleic Acids Res*, 1993. №21(25) p. 5803-5816.
- 2) Coletta A. Low-complexity regions within protein sequences have position-dependent roles // *BMC Syst Biol*, 2010. №4 p. 43.
- 3) Kumari B. Low complexity and disordered regions of proteins have different structural and amino acid preferences. // *Mol Biosyst*, 2015. №11(2) p. 585-94.
- 4) Ozdilek B.A. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. // *Nucleic Acids Res*, 2017. №45(13) p. 7984-7996.

- 5) Radó-Trilla N. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins // BMC Evol Biol, 2012. №12 p. 155.
- 6) Steinert P.M. Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricroins and single-stranded RNA binding proteins. // Int J Biol Macromol, 1991. №13(3) p. 130-9.
- 7) Thandapani P. Defining the RGG/RG motif. // Mol Cell, 2013. №50(5) p. 613-23.