

УДК 577.052.6,577.21,577.112.4,577.113

САЙТЫ ФОСФОРИЛИРОВАНИЯ ТЯГОТЕЮТ К АЛЬТЕРНАТИВНО СПЛАЙСИРУЕМЫМ ОБЛАСТЯМ БЕЛКОВ

© 2009 г. Е. Ж. Курмангалиев^{1,2}, М. С. Гельфанд^{1,3,*}

¹Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, 127994

²Национальный центр биотехнологии Республики Казахстан, Астана, 010000

³Факультет биоинженерии и биоинформатики Московского государственного университета им. М.В. Ломоносова, Москва, 119992

Поступила в редакцию 19.08.2008 г.

Принята к печати 01.10.2008 г.

Ключевые слова: альтернативный сплайсинг, сайты фосфорилирования.

ALTERNATIVE SPLICING TENDS TO INVOLVE PHOSPHORYLATION SITES, by Y. Z. Kurmangaliyev^{1,2}, M. S. Gelfand^{1,3,*} (¹Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia, *e-mail: gelfand@iitp.ru; ²National Center for Biotechnology of the Republic of Kazakhstan, Astana, 010000 Kazakhstan; ³Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119992 Russia)

Key words: alternative splicing, phosphorylation sites.

Фосфорилирование – один из наиболее важных и распространенных типов посттрансляционных модификаций белков. Последние достижения в методах масс-спектрометрии позволяют получать массовые данные о фосфорилировании белков. По некоторым оценкам, как минимум, одна треть всех белков протеома человека фосфорилированы [1, 2].

Недавние эксперименты, результаты которых доступны в базе данных PHOSIDA, позволили идентифицировать более 6000 сайтов фосфорилирования в протеоме человека [3, 4]. На предсказанных вторичных структурах белков показано, что фосфосерины и фосфотреонины предпочитают располагаться в петлях, в то время как фосфотирозины – в областях с упорядоченной вторичной структурой. Также показано, что сайты фосфорилирования более консервативны, чем петли в целом, и чем нефосфорилированные аминокислотные остатки [4].

С другой стороны, альтернативный сплайсинг существенно влияет на структуру и функцию белков [5]. Кроме того, альтернативный сплайсинг часто затрагивает функциональные сайты [6] и домены, участвующие в процессах внутриклеточной сигнализации [7, 8].

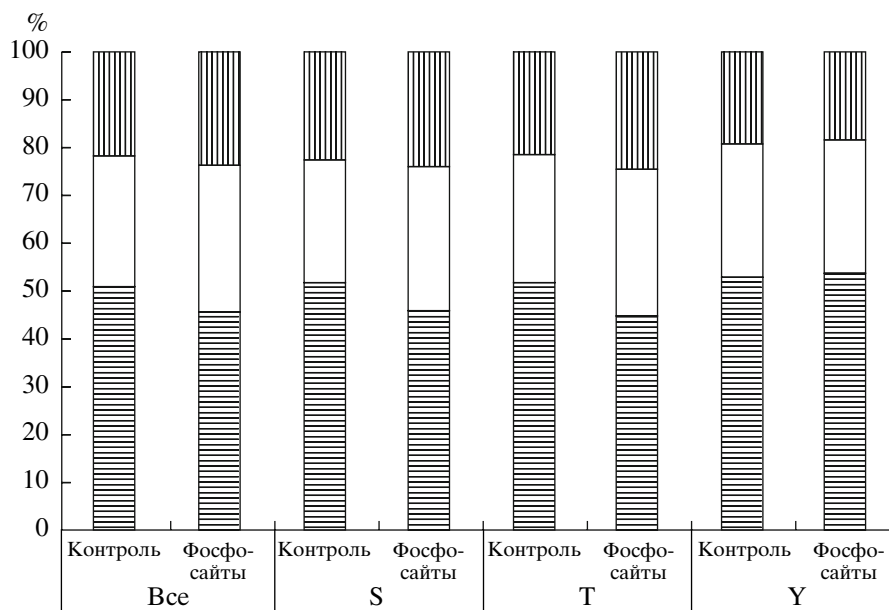
Поэтому задачей данного исследования стало изучение взаимосвязи между фосфорилированием и альтернативным сплайсингом с использованием ставших недавно доступными массовых данных.

* Эл. почта: gelfand@iitp.ru

Набор сайтов фосфорилирования получен из фосфопептидов, идентифицированных в клетках HeLa человека и собранных в базе данных PHOSIDA [3, 4]. На следующем этапе мы определили расположение данных фосфосайтов в геноме путем картирования фосфопептидов на белках из базы данных EDAS [9]. Картирование проводили поиском точного совпадения последовательности фосфопептида в любой изоформе для всех генов. Далее использовали лишь фосфопептиды и фосфосайты с однозначным геномным расположением. Конечная выборка состояла из 7584 фосфосайтов в 1907 генах человека; из них 5990 фосфосеринов, 1312 фосфотреонинов и 282 фосфотирозина. В качестве контрольной выборки использовали нефосфорилированные остатки серина (150340 остатков), треонина (87677) и тирозина (37899) наиболее длинных изоформ белков.

Данные по альтернативному сплайсингу, в том числе “включаемость” кодонов, косвенно указывающая на частоту включения кодируемой аминокислоты в альтернативных изоформах белка, также получены из базы данных EDAS [9]. “Включаемость” определена как доля последовательностей EST, содержащих кодон, среди всех последовательностей EST, покрывающих данную область гена.

Аминокислотные остатки классифицировали как конститутивные (принадлежащие к конститутивным сегментам, включаемость 100%) и альтернативно сплайсируемые (принадлежащие к альтернативно сплайсируемым сегментам белков,



Распределение фосфорилированных и нефосфорилированных аминокислотных остатков: в конститутивных (горизонтальный штрих), неопределенных (белое) и альтернативно сплайсируемых (вертикальный штрих) участках для серинов (S), треонинов (T), тирозинов (Y) и всех этих остатков суммарно (все).

включаемость менее 99%). Около 30% всех остатков, включая сайты фосфорилирования, имели долю включения 99%. Мы их отнесли к группе неопределенных, так как это могло быть как результатом альтернативного сплайсинга, так и следствием ошибок сплайсинга или экспериментальным артефактом.

Анализ данных показал, что сайты фосфорилирования имеют тенденцию находиться в альтернативно сплайсируемых областях белков (рисунок). Во всех случаях эта тенденция слаба, но статистически достоверна. В 24% случаев сайты фосфорилирования альтернативно сплайсируемы, а в 46% случаев – конститутивны. В то же время, для нефосфорилированных остатков эти величины равны 21 и 51% соответственно (тест χ^2 , $P = 0$).

Аналогично, при отдельном рассмотрении фосфосеринов, 24% находилось в альтернативно сплайсируемых областях и 45% в конститутивных (тест χ^2 , $P = 2 \times 10^{-8}$). Сходные данные получены в случае фосфотреонинов – 25 и 44% соответственно (тест χ^2 , $P = 1 \times 10^{-4}$). В то же время, для относительно небольшой выборки фосфотирозинов никакой значимой разницы не обнаружено.

Тем самым, полученные данные согласуются с ранними наблюдениями о корреляции альтернативного сплайсинга и функционально важных участков белков [6]. Одним из возможных объяснений полученного результата может являться то, что как и для сайтов фосфорилирования, так и для альтернативно сплайсируемых областей показана тенденция к расположению на поверхности белков [10] и в неструктурированных областях [11]. Эта гипотеза может

быть проверена после того, как станут доступны пространственные структуры достаточного количества фосфорилируемых белков.

Мы благодарны Д. Фришману за идею настоящего исследования, а также Р. Нуртдинову за помощь в работе с базой данных EDAS.

Работа частично поддержана программой "Молекулярная и клеточная биология" РАН и Российским фондом фундаментальных исследований (07-04-00343 и 08-04-90900).

СПИСОК ЛИТЕРАТУРЫ

1. Cohen P. 2001. The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur. J. Biochem.* **268**, 5001–5010.
2. Ptacek J., Snyder M. 2006. Charging it up: global analysis of protein phosphorylation. *Trends Genet.* **22**, 545–554.
3. Olsen J.V., Blagoev B., Gnad F., Macek B., Kumar C., Mortensen P., Mann M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell.* **127**, 635–648.
4. Gnad F., Ren S., Cox J., Olsen J.V., Macek B., Oroshi M., Mann M. 2007. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8**, R250.
5. Artamonova I.I., Gelfand M.S. 2007. Comparative genomics and evolution of alternative splicing: the pessimists' science. *Chem. Rev.* **107**, 3407–3430.
6. Kriventseva E.V., Koch I., Apweiler R., Vingron M., Bork P., Gelfand M.S., Sunyaev S. 2003. Increase of

- functional diversity by alternative splicing. *Trends Genet.* **19**, 124–128.
7. Liu S., Altman R.B. 2003. Large scale study of protein domain distribution in the context of alternative splicing. *Nucleic Acids Res.* **31**, 4828–4835.
 8. Yura K., Shionyu M., Hagino K., Hijikata A., Hirashima Y., Nakahara T., Eguchi T., Shinoda K., Yamaguchi A., Takahashi K., Itoh T., Imanishi T., Gojobori T., Go M. 2006. Alternative splicing in human transcriptome: functional and structural influence on proteins. *Gene.* **380**, 63–71.
 9. Nurtdinov R.N., Neverov A.D., Mal'ko D.B., Kosmodem'ianskii I.A., Ermakova E.O., Ramenskii V.E., Mironov A.A., Gel'fand M.S. 2006. EDAS, databases of alternatively spliced human genes. *Biofizika.* **51**, 589–592.
 10. Wang P., Yan B., Guo J.T., Hicks C., Xu Y. 2005. Structural genomics analysis of alternative splicing and application to isoform structure modeling. *Proc. Natl. Acad. Sci. USA.* **102**, 18920–18925.
 11. Romero P.R., Zaidi S., Fang Y.Y., Uversky V.N., Radivojac P., Oldfield C.J., Cortese M.S., Sickmeier M., LeGall T., Obradovic Z., Dunker A.K. 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. USA.* **103**, 8390–8395.